NASA CONTRACTOR REPORT

NASA CR-18[83]

NASA CR-1883

AN EVALUATION OF METHODS FOR SCALING AIRCRAFT NOISE PERCEPTION

by J. B. Ollerhead

Prepared by
WYLE LABORATORIES
Hampton, Va. 23366
for Langley Research Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • OCTOBER 1971

1. Report No. NASA CR-1883	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle AN EVALUATION OF METHODS FOR SCALING AIRCRAFT NOISE PERCEPTION		5. Report Date November 1971
		6. Performing Organization Code
7. Author(s) J. B. Ollerhead		8. Performing Organization Report No.
9. Performing Organization Name and Address Wyle Laboratories Hampton, Virginia		10. Work Unit No.
		11. Contract or Grant No. NAS 1-9257
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, D.C. 20546		13. Type of Report and Period Covered Contractor Report
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

Following a review of previous research, an extensive experiment was undertaken to assess the practical differences between numerous alternative methods for calculating the perceived level of aircraft flyover sound. One hundred and twenty recorded sounds, including jets, turboprops, piston engined aircraft and helicopters were rated by a panel of subjects in a paired comparison test. The results were analyzed to evaluate a number of noise rating procedures in terms of their ability to accurately estimate both relative and "absolute" perceived noise levels over a wider dynamic range than has generally been used in previous experiments.

It was found that the "complex" procedures developed by Stevens, Zwicker and Kryter are superior to other scales, particularly when integrated to include a signal duration allowance. The main advantage of these methods over the more convenient weighted sound pressure level scales lies in their ability to cope with signals over a wide range of bandwidth. However, Stevens' loudness level scale and the perceived noise level scale both overestimate the growth of perceived level with intensity because of an apparent deficiency in the band level summation rule, which is common to both. A simple correction is proposed which will enable these scales to properly account for the experimental observations.

17. Key Words (Suggested by Author(s)) Aircraft noise, Perceived noise level		18. Distribution Statement Unclassified – Unlimited	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 149	22. Price* $3.00

*For sale by the National Technical Information Service, Springfield, Virginia 22151

CONTENTS

TABLES

S_t	total perceived magnitude
SPL	Sound Pressure Level
T_c	effective (integrated) duration
Δ	mean value of $x_i - y_i'$

ACKNOWLEDGEMENTS

The author wishes to thank B. D. Adcock for his major contribution to the planning, execution and analysis of the experimental program. He is also indebted to K. M. Eldred for many helpful discussions and a critical review of the final report. Significant contributions to the study were also made by D. M. Lister and A. C. Jolly (Computer Programming), E. N. Esslinger and C. Price (Instrumentation) and J. Matzkiw and R. Brye (Test Directors).

AN EVALUATION OF METHODS FOR SCALING AIRCRAFT NOISE PERCEPTION

By J. B. Ollerhead
Wyle Laboratories
Hampton, Virginia

SUMMARY

Following a review of previous research, an extensive experiment was undertaken to assess the practical differences between numerous alternative methods for calculating the perceived level of aircraft flyover sound. One hundred and twenty recorded sounds, including jets, turboprops, piston engined aircraft and helicopters were rated by a panel of subjects in a paired comparison test. The results were analyzed to evaluate a number of noise rating procedures in terms of their ability to accurately estimate both relative and "absolute" perceived noise levels over a wider dynamic range than has generally been used in previous experiments.

It was found that the "complex" procedures developed by Stevens, Zwicker and Kryter are superior to other scales, particularly when integrated to include a signal duration allowance. The main advantage of these methods over the more convenient weighted sound pressure level scales lies in their ability to cope with signals over a wide range of bandwidth. However, Stevens' loudness level scale and the perceived noise level scale both overestimate the growth of perceived level with intensity because of an apparent deficiency in the band level summation rule, which is common to both. A simple correction is proposed which will enable these scales to properly account for the experimental observations.

The better scales performed consistently for the jet and piston engine sounds, but showed deficiencies for application to the turboprop and helicopter sounds. It is believed that improvements to the tone correction might remedy some of these deficiencies, but that the perception of low frequency harmonic sound needs further study.

It is recommended that the search for an improved sound level meter scale be continued, not as a replacement for the more complex perceived level procedures, but to supplement them, particularly for monitoring purposes. Despite deficiencies which cannot be overcome by refined weighting circuits, it is clear that the weighted sound pressure level provides a very powerful scale for comparing the sounds of aircraft.

1.0 INTRODUCTION

Concern about aircraft noise intrusion has grown at about the same rapid pace as progress in commercial aviation. The need for swift economical transportation between population centers has brought large segments of the community within earshot of modern airports, and experience has made it clear that substantial aircraft noise reductions must pave the way for further progress.

The complexity of the airport noise problem is largely related to the high cost of aircraft noise suppression. If inexpensive mufflers could be fitted to aircraft the problem would have been solved long ago. Instead, a compromise must be sought which is a very delicate balance between comfort, convenience and cost. Unfortunately, the cost of aircraft noise reduction is extremely sensitive to the noise parameter. A decibel one way or the other can have a large effect on aircraft economics, and it is for this reason that a great deal of importance has been attached to the problem of specifying and measuring noise levels.

In many planning and regulatory aspects of aircraft noise control, it is necessary to define noise in terms which are related to human evaluation. It has long been known that humans react to noise in a more complex manner than accounted for by the simple measure of overall sound pressure level. Unfortunately, the search for a more suitable scale, upon which the subjective magnitude of aircraft sound can be accurately related to physical measurements, has proved to be surprisingly difficult. After two decades of research, there remains considerable confusion about the relationships between a multitude of alternative solutions and the problem seems far from solved.

The main difficulty is that people vary considerably in their response to noise intrusion. In the "real situation," where people are bothered by aircraft noise, many factors other than the physical characteristics of the intruding sound contribute to the disturbance, including the ambient noise levels, hearing acuity, activity, how frequently the intrusion occurs, and so on. Also, it is practically impossible to quantitatively study the problem within a realistic environment since the very process of making the necessary observations has a significant effect upon subjective reaction.

However, despite these complexities, one fact seems self-evident -- the greater the noise intrusion, the more disturbed people are likely to be. Indeed, this is one clear finding that does emerge from community noise surveys. The Heathrow study of 1961 (ref. 1) resulted in a recommendation for the use of a "Noise and Number Index" to quantify community noise exposure. This index very simply stated that community annoyance is related to both the average noise level experienced during each aircraft flyby and the number of flybys which occur during a given period. Similar concepts are now in use throughout the world.

The present study is concerned with the definition of the noise level associated with an aircraft flyby. The subjective magnitude of a sound has been described by a variety of terms, including loudness, noisiness, and intrusiveness, and measurement scales have been labeled loudness level, perceived noise level, annoyance level, and so on. To avoid confusion herein, the term "perceived level," recently suggested by Stevens (ref. 2) is adopted to maintain some generality and to avoid the implication of meaningful differences between the various terms. The term will frequently be further qualified by the words "calculated," to indicate that the level was obtained through the analysis of objective measurement, and "judged," which refers to a subjectively measured quantity. A distinction between perceived level and perceived magnitude should be noted. The latter, corresponding to loudness, noisiness, etc., is numerically proportional to the subjective quantity, whereas the term perceived level relates to the value of the quantity upon a logarithmic (decibel) scale of measurement.

Factors which contribute to the perceived level of sound have been studied extensively over a period of more than forty years. The pioneering work of Steinberg (ref. 3), Fletcher and Munson (ref. 4), revealed that the ear is very frequency sensitive, and that perceived level can vary by more than 50 dB for pure tones at various frequencies in range. Fletcher (ref. 5), and later Stevens (ref. 6) and Zwicker (ref. 7) hypothesized the process by which energy in different parts of the spectrum add their contributions to perceived level and defined rules for its mathematical simulation. Further research, involving both synthetic (laboratory generated) sounds and real recordings, has indicated the possibly important roles of time variations in the sounds and the presence of intense concentrations of energy in very narrow segments of their spectra.

Unfortunately, again because of inherent human variability and the difficulty of subjective measurement, these studies have led to little concensus of opinion. For example, more than eighteen different studies of the dependence of perceived level upon frequency have been made and hardly any two agree. Largely on the basis of these various frequency functions, at least twenty different procedures for rating noise have been proposed, many of which have been adopted and are in use for a variety of purposes. The development of advanced high speed computers and instrumentation for acoustic analysis has made previously undreamed of complexity possible and there is now hardly any limit to the number of independent factors which can be included in a fairly practical noise rating procedure.

Many studies have been performed to evaluate the relative merits of different procedures and to attempt to develop an 'optimum' one. This quest for improvement is often questioned on the grounds that, due to the large measurement scatter encountered in subjective experimentation, it would be difficult to identify a 'perfect' scale even if one existed. Also, the need for a perfect scale must be related to its intended purpose. In the Heathrow study (ref. 1), for example, the highest correlation which could be achieved between community annoyance and the best indices of noise was a coefficient of less than 0.5. It is unlikely that any refinement of the perceived level scale could result in a substantial improvement in this situation where many factors in addition to

the noise level contribute to the problem. On the other hand, any scale which forms the basis for noise regulation is susceptible to strong criticism and must, therefore, have demonstrated validity and the highest possible accuracy.

A review of research into the subjective evaluation of aircraft noise shows that although many comparisons of different noise rating procedures have been based upon studies of a wide variety of acoustic stimuli, at no time has an extensive examination of their applicability to aircraft noise been made. Typical experiments have involved perhaps a dozen different aircraft sounds which is, of course, a small statistical sample. Attempts to utilize accumulative evidence of a number of such studies, as recently attempted by Kryter (ref. 8), can be thwarted by the different experimental techniques of different investigators.

The present study, therefore, was initiated with the intention of performing a large scale subjective experiment to obtain a hitherto unavailable quantity of self-consistent data. This data, comprising more than one hundred judged perceived levels for a large variety of recorded aircraft flyover sounds, would then provide a basis for an adequate statistical comparison of the various recommended perceived level scales. Consistent deficiencies in the scales might be explained and supporting experiments could be performed to develop recommendations for refinements, revisions or additions. A particular objective was to investigate the possible effects of aircraft motion which an earlier study (ref. 9) had suggested might be of importance. This was done through a subsidiary experiment involving sounds which were synthesized to exhibit the acoustic effects of the relative motion of a simple source of sound. The main experimental program was performed in parallel with a Federal Aviation Administration sponsored investigation of the noise of STOL aircraft (ref. 10). The two experiments were performed in an almost identical manner so that the two sets of data could be combined for the above purposes. The combined sample includes results for recorded sounds of one hundred and twenty aircraft, including piston engined, turboprop, turbojet, and turbofan powered aircraft and helicopters with gross weights between 2300 and 315,000 pounds. The sounds ranged in (peak) level from 84 to 115 dB*, and had 10 dB-down durations between 1 and 27 seconds. Thus, the experiment provided a test of the noise rating methods which is unprecedented in scale and severity.

The next section of this report is devoted to a review of previous studies of subjective response to aircraft noise. Emphasis has been placed upon investigations which were specifically addressed at the aircraft noise problem, but generally the discussion supplements a review of more basic research into the perceived level of complex sounds performed by Kryter (ref. 11) in 1966.

---

* Throughout this report, sound pressure levels are expressed in dB relative to 2×10^{-5} N/M^2 unless otherwise stated.

2.0 REVIEW OF PREVIOUS RESEARCH INTO THE PERCEIVED LEVEL OF AIRCRAFT NOISE

2.1 Basic Factors and Definitions in Perception of Complex Noise

Sound can be measured in terms of its level or intensity, its frequency distribution of energy, its spectral character (for example, whether it is harmonic, random, impulsive, etc.), and the variation of these quantities with time. For the purpose of subjective evaluation, it is necessary to know how each of these factors contributes to perceived level.

2.1.1 Frequency and intensity. - The earliest studies of perceived magnitude were concerned with the variations of the loudness of pure tones with frequency. Fletcher and Munson (ref. 4) in 1933 produced the first set of equal loudness contours which eventually found their way into the American Standards, and which are still widely applied through the approximations employed by the A, B, and C weighting networks of standard sound level meters (ref. 12). Since that date at least eighteen further studies have been performed, which have been addressed at both tones and bands of noise, different listening conditions (free field, diffuse and earphone presentation), and different descriptions of perceived magnitude, in particular loudness and noisiness. Worthy of particular note is the extensive study made by Robinson and Dadson (ref. 13) in 1956. However, the fact that each of these studies produced at least one and sometimes several new and different equal magnitude contours, as illustrated in figures 1 and 2, is probably the first and most serious source of confusion. The differences can be attributed to many sources of variation, the type and quality of sounds utilized, the particular group of human subjects, the experimental technique, the listening conditions, the accuracy of measurement, the method of analysis, the instructions given to the subjects, the preconceptions of the investigator, and so on. While many of the observed differences might indeed be attributable to the particular influences under study, the fact that different measurements of purportedly the same quantity differ by equal margins does tend to obscure the sources of variability. It is for this reason that Robinson (ref. 14) and more recently Stevens (ref. 2), recommended that the most reasonable approach is to take an average of all appropriate results. However, although Stevens (ref. 2) has now defined a particular set of averaged curves, the ones in most general use are based upon measurements by Fletcher and Munson (ref. 4), Stevens (ref. 6), and Kryter and Pearsons (ref. 15).

The variation of perceived magnitude with signal intensity has been studied by many investigators and for many different sounds, including tones and bands of noise. It has been found that perceived magnitude is related to signal intensity by a power law of the type

$$S \sim (E - E_o)^k \qquad (1)$$

where E_o is a threshold intensity, and k is a constant around 0.3. At intensities well away

from the threshold, this, in fact, corresponds to a doubling of perceived magnitude each time the stimulus level increases by 10 dB. Although this 10 dB-per-doubling value is much used and much quoted, small but distinct variations of the constant k have been noted by Zwicker (ref. 7), Robinson (ref. 14), and Stevens (ref. 16). Some of these differences are attributable to experimental procedure. For example, if a higher frequency reference sound is used to measure the growth of perceived magnitude of a lower frequency sound, the exponent k is determined to be variable. Figure 3 shows this effect in curves derived from loudness matches between different frequency tones, which indicate that the growth of loudness varies with both frequency and level over the entire sound pressure level range of practical interest. Note that the slope of the curves in figure 3, which asymptotes to a value of 0.85 at the higher levels, indicates that the perceived level of the lower frequency sounds increases at a higher rate. Regardless of the choice of reference sound, Robinson (ref. 14), has shown that the growth of perceived magnitude varies with sound pressure level, generating a distinct "mid-level bulge," as Stevens calls it. This growth function, in dB per two-fold perceived magnitude change, will be seen in figure 11.

2.1.2 The summation of perceived magnitude. - Although the perceived magnitude of simple acoustic stimuli such as tones and narrow bands of noise can be specified with relative ease, the manner in which the individual magnitudes of a number of such components add, in the case of a more complex sound, does not lend itself to simple treatment.

Although many approaches have been suggested, that developed by Zwicker in 1958 (ref. 7) appears to be the most soundly constituted calculation procedure presently available. Because of the importance of many of the basic concepts incorporated, Zwicker's method will be reviewed in some detail here.

Zwicker's Loudness Level Computation

The approach is developed around the critical band ("frequenzgruppe") hypothesis which relates the frequency analysis function of the hearing mechanism to certain linear divisions of the basilar membrane. The existence of "critical bands" was first postulated by Fletcher in 1940 (ref. 5) and many studies since, especially by Zwicker, Flottorp and Stevens (ref. 17), have revealed that the concept explains many of the basic properties of the auditory mechanism, including frequency discrimination, loudness summation, and masking. Very simply, it is hypothesized that any given frequency or narrow band of frequencies excites a particular segment of the basilar membrane. The signals from different segments are measured by the organ of Corti, which is attached to the membrane, and transmitted to the brain for integration. The entire length of the membrane is about 31mm, and this is divided into 24 "Kopplungsbreiten" or critical widths of approximately 1.3mm, each of which corresponds to a particular critical band. Figure 4, based on data published by Zwicker in 1961 (ref. 18), shows the frequency dimensions of the critical bands which may be thought of as the built-in filter bank of the hearing system.

The role of the critical bands in the summation of perceived magnitude is such that if the acoustic energy is confined within a critical bandwidth, the perceived level is roughly proportional to sound pressure level. However, if the energy is distributed over a wider frequency range, the perceived magnitudes of different critical bands do not add according to such a simple rule. The reason for this is that energy in one band can mask the sound in adjacent bands. In particular, Zwicker noted that as the level of a tone or a narrow band of noise is raised, an increasingly greater length of the organ of Corti is stimulated. Thus, if a second tone is added at a different frequency, it attempts to excite nerve fibers which are already responding to the first stimulus, and do not, therefore, respond linearly to the additional excitation. This masking effect is illustrated in figure 5 which shows the "upwards and outward" spreading of the masked threshold caused by an increase in level of a narrow band of noise centered at 1200 Hz. Zwicker interpreted this effect as a demonstration that the loudness of a tone or narrow band of noise is composed of several "partial loudnesses" so that the whole could be obtained by effectively integrating the area between the curves of figure 5 and the threshold of audibility. To do this he first had to correct the ordinate to specific loudness (which is analogous to power spectral density) using the assumed relationship

$$\frac{dS}{df} \sim \left(\frac{E}{E_o}\right)^k \tag{2}$$

where E_o is the threshold level (note that this equation is similar in form to equation (1)). The result of this step for the 1200 Hz tone with a level of 100 dB is shown in figure 6 as a curve of loudness density (in loudness per critical band) plotted against band number. To simplify the curve for the purposes of practical application, Zwicker ignored the rather small downward spread of masking and defined the approximate curve also shown in figure 6. The exponent k in equation (2) takes the value 0.25 so that the area under the specific loudness curve doubles each time the signal sound pressure increases by 10 dB. This will be discussed further below.

Up to this point consideration had been given to the signal level, its bandwidth, the threshold of audibility and the effect of masking. A final step involved the transfer function which translates atmospheric pressure fluctuations into effective amplitude fluctuations at the "doorway" to the organ of Corti, the oval window. This may be seen superimposed on the integrated diagram of figure 7 which is Zwicker's original loudness calculation diagram. Each of the solid curves corresponds to a critical band sound pressure level and accounts for the above-mentioned transfer function for free field sound. The loudness corresponding to the energy in any band is proportional to the area in that band under the appropriate sound pressure level curve, plus the area under the upward masking "sideband." However, this latter area must be subtracted from that measured under any higher bands into which it encroaches.

In order to make his procedure tractable for practical purposes, Zwicker (ref. 19) developed a set of calculation diagrams for use with 1/3-octave band level data obtainable from conventional analysis equipment. Several charts are presented for both free field and diffuse listening conditions and for high and low levels. An example is presented in figure 8 which again is for free field exposure and has been used to demonstrate its application to a narrow-band sound. The total loudness, S_t, in any case, is proportional to the area under the composite curve. This magnitude is then converted to loudness level using the relationship for a "standard" 1000 Hz tone:

$$LL_Z = 33.3 \log_{10} S_t + 40 \tag{3}$$

where LL_Z is the loudness level in phons and S_t is the calculated total loudness in sones (Note that this corresponds to an exponent of $k = 0.3$ in equation (1)). The constant 40 arises because one sone is defined as the loudness of a tone of frequency 1000 Hz and sound pressure level 40 dB. The loudness level of any sound, in phons, is thus numerically equal to the sound pressure level of the 1000 Hz tone which is equally loud.

The above conversion is indicated by the phon scales at the sides of figure 8. It may be noted that although length on this scale doubles for each 10 dB increment, the same is not true of the scales in the main 1/3-octave chart. This reflects the fact that for each 10 dB increment of sound pressure level, it is the total area under the specific loudness curve for any single band which doubles. This can be seen in the series of curves drawn for 10 dB increments of the 1/3-octave band of noise centered at 1000 Hz. In accordance with experimental measurements of the growth of loudness at that frequency, it will later be seen that this fact assumes considerable practical importance in the present study.

Stevens' Loudness Level Computation

A second technique for the calculation of the loudness of complex noise has been developed by Stevens (references 6 and 20). Because this method is considerably simpler than Zwicker's, it tends to be used more widely although both methods have been adopted by the American Standards Association (ref. 21).

In the first place, Stevens confined his attention to sound measurements made in conventional octave or 1/3-octave bandwidths. Secondly, like Zwicker, he invoked the power function (in limited regions of the frequency-level plane) given by equation (1) to define independently the loudness of each band of noise in units of 'sones'. However, he retained the exponent $k = 0.3$ and ignored the "energy spreading" effect at this stage. Instead, he accounted for this in a rule for summing the individual loudnesses. The rule, very simply, assumes that the loudest band contributes fully to the total loudness. However, because of the masking effect, the contributions of all other bands are inhibited to

a fraction of the loudness they would exhibit if heard alone. The summation rule is, simply

$$S_t = S_m + F(\sum S - S_m) \tag{4}$$

where S_t is the total loudness, S_m is the loudness of the loudest band, and $\sum S$ is the sum of all band loudnesses. The factor F was determined experimentally to be 0.3 for octave bands, or 0.15 for 1/3 octave bands of noise. Finally, the total loudness, S_t, is converted to a loudness level LL_S through the use of equation (3).

Stevens observed departures from the power function (1) at low levels and low frequencies and since, in any case, the constant E_o has to be determined empirically, complete curves and tables for the intensity loudness relations were presented in reference (6). In 1961, Stevens (ref. 22) revised his original loudness procedure and presented a version labeled Mk VI. This revision was made for two reasons: (i) to approximate the earlier equal loudness contours by straight lines (on a log-log plot) for the convenience of computation, and (ii) to vary the spacing between contours to accommodate the "mid-level bulge" identified by Robinson. It was claimed that the revised procedure agreed better with available loudness measurements.

Very recently Stevens (ref. 2) progressed to his Mk VII procedure which attempts to include state-of-the-art knowledge into an extensively revised scheme. This includes: (i) a new basic reference sound (a tone or narrow band of noise near 3000 Hz), which helps to avoid some non-linearities associated with the standard reference at 1000 Hz, (ii) a new set of equal loudness contours extended down to a frequency of 1 Hz and averaged over the results of many different investigators, (iii) the use of an F-factor which varies as a function of level to account for the bulge previously introduced through variable spacing of the loudness contours, (iv) a change in the exponent k to the value 0.33 so that loudness doubles every 9 dB and finally, (v) the rejection of the terms loudness and loudness level in favor of perceived magnitude and perceived level. The implications of these changes will be discussed below.

A revised version of Stevens' loudness level procedure, which has gained widespread popularity nationally and internationally, particularly in aviation circles, is the Perceived Noise Level concept due to Kryter (ref. 23). In its original form, this differed from Stevens' loudness level by the substitution of a set of equal "noisiness" contours for the equal loudness curves and the adoption of the attribute "perceived noise" as a more appropriate evaluator than loudness*. This was based on the finding that this

*In this procedure the equivalent of the 'sone' unit is the 'noy' where one noy is defined as the noisiness of an octave band of random noise centered at 1000 Hz with a sound pressure level of 40 dB.

new procedure performed more consistently for aircraft noise than did the Mk VI method. There has since been considerable speculation as to whether or not the difference between loudness and noisiness is real, a distinction which Stevens (ref. 2) has now chosen to avoid by replacing both quantities by the single term "perceived magnitude".

Since 1959, the perceived noise level procedure has undergone a large number of revisions, refinements and extensions which were reviewed in detail by Kryter in 1968 (ref. 24). These modifications account for temporal and spectral complexities not considered in the basic procedures and improve correlation between calculated and measured levels for aircraft flyover noise. These modifications will be described in section 2.2.

Weighted Sound Pressure Level Estimates

A third basic method used for estimating perceived level is that of simple energy summation in the weighted sound pressure level. This is obtained by passing the acoustic signal through a filter whose frequency response corresponds to the inverse of an equal perceived magnitude contour. This approach is extremely popular because of its practical simplicity and numerous appropriate filter functions have been recommended. However, only three remain available on the standard sound level meter -- the A, B, and C scales, which approximate the inverse of the Fletcher and Munson 40, 70 and 100 phon contours, respectively. An improved network to be called the D-scale, is presently under consideration, but a concensus of opinion on an optimum shape is proving difficult to achieve.

Comparison of the Three Procedures

Since each of the three basic summation procedures described above is supposed to estimate the perceived level of any broadband sound, it is of interest to examine how the three methods compare with each other in typical applications.

In order to avoid the influence of different frequency weighting functions which may tend to obscure the magnitude summation effects we are trying to examine, the concept of "uniformly distributed noise" is introduced. This is a broadband noise having a spectrum such that each individual band, if present alone, would independently yield the same perceived level on the scale in question. Thus, each calculation method has its own particular noise spectrum; the particular shapes of these spectra are of no direct interest here.

The first point of interest is the manner in which the perceived magnitudes of N bands of noise add together to give an overall perceived level. For the weighted sound pressure levels, the sum is simply given by

$$L_W = PL_{1/3} + 10 \log_{10} N \qquad (5)$$

where $PL_{1/3}$ is the perceived level of each band. Thus, the total level increases by 3 dB each time the number of contributing bands is doubled. This, of course, is simply the energy summation principle.

Equating S_m and S for our uniformly distributed noise in Stevens' summation rule (Equation (4)), we see that

$$LL_S = 33.3 \log_{10} S \left\{1 + F(N-1)\right\} \tag{6}$$

The quantity $33.3 \log_{10} S$ is equal to the band perceived level $PL_{1/3}$ so that we can write for 1/3-octave bands of noise ($F = 0.15$):

$$LL_S = PL_{1/3} + 33.3 \log_{10} (0.85 + 0.15\, N) \tag{7}$$

When N is very large, the total level increases by 10 dB each time the number of admitted bands is doubled. However, the increment is less for a more realistic number of bands as shown in figure 9, where the curve given by equation (7) is seen to cross the 3 dB per doubling line of equation (5) at between $N = 5$ and $N = 6$.

A similar analysis may be applied to Zwicker's rule. An analysis of the "specific loudness" charts similar to the one shown in figure 8 reveals that for any given sound pressure level, the proportion of the loudness of any simple 1/3-octave band of noise confined to the sideband masking envelope is roughly equal for all bands. Thus, the proportion actually confined between the 1/3-octave band frequency limits f_{1_n} and f_{2_n} can be expressed for one band as follows:

$$\frac{\int_{f_{1_n}}^{f_{2_n}} \frac{dS}{df}\, df}{\int_{0}^{\infty} \frac{dS}{df}\, df} = F \tag{8}$$

where the subscript n denotes the n-th band. It can be seen upon inspection of figure 8 that the sideband associated with each additional and adjacent band is masked when it is added to the first, so that the total loudness for N bands is approximately

$$S_t = S \left\{1 + F(N-1)\right\} \tag{9}$$

This relationship is, of course, identical to the equation for Stevens' summation principle defined above.

Approximate average values for F have been estimated from Zwicker's charts (ref. 19) as a function level and are shown in figure 10 to steadily decrease from around 0.6 at low sound pressure levels to a little more than 0.2 at 110 dB. Thus, at all levels the

factor is greater than the value of 0.15 originally suggested by Stevens for 1/3-octave band summations. However, attention is drawn to Stevens' revised and variable F-function used in the Mk VII procedure (ref. 2), which is included for comparison in figure 10. In view of the totally different derivations of the two curves, they are remarkably similar at levels above 50 dB. They do, in fact, coincide at 110 dB. The reason for the disparity below 50 dB is related to the fact that Stevens' curve takes account of departures from the power law (equation (2)) at low levels.

The perceived level summation increment corresponding to F= 0.3 is included for comparison in figure 9. This corresponds to the Zwicker case at band levels around 80 dB. However, variation of F between 0.4 and 0.2 at band levels above 50 dB can cause the curve to vary over a total range of 10 dB about its illustrated position. Similar variations occur in the corresponding curve for Stevens' Mk VII summation rule due to the variable F-factor. However, the difference between the curve corresponding to the lowest value F= 0.19 which occurs around 80 dB (see figure 10) and the Mk VI curve for F= 0.15 is very small due to different sone/phon conversion factors of 33.3 (Mk VI) and 30 (Mk VII).

Thus, we see that at band levels around 80 dB, the three basic procedures give different weightings to the perceived level increment caused by adding further bands of noise. For more than 6 equal magnitude bands, the energy summation principle (sound pressure level scales) gives the smallest increment, followed by the Stevens' method and Zwicker's method. However, the differences do vary with level, and, of course, with spectrum shape. Further, the absolute differences between the Zwicker and Stevens curves are in practice reduced by other procedural differences in the level computations.

Turning now to the growth of perceived level with intensity, reference is again made to the "uniformly distributed noise" concept. As the sound pressure level of this noise is raised, the perceived level estimated on any weighted sound pressure level scale increases an equal amount, by definition. Thus, the growth curve is a straight line with a unit slope as shown in figure 11. The same curve also applies to the Perceived Noise Level (PNL) scale at levels above 50 dB since it is based on a uniform 10 dB per doubling of perceived magnitude.

The same is not true of Zwicker's scale, however. Although the "10 dB per doubling" approximation has been adopted by Zwicker, it only remains true in the conversion from sound pressure level to loudness for individual 1/3-octave bands of noise where the full energy spreading effect contributed to loudness (i.e. the area under the curve doubles). Thus, for single bands of noise, perceived level increases at the same rate as sound pressure level (at levels greater than approximately 60 dB). However, in the case of multiple adjacent bands of noise, the "sideband loudness" is masked to an extent which represents an increasing proportion of the total as level increases. Thus, it turns out that for all bands other than the first, the loudness is given by the power law

$$S \sim E^{0.25} \tag{10}$$

This rule remains approximately true for the entire spectrum if it contains many bands of noise where the spread of masking contributed little loudness to the total. However, in the final conversion back from loudness level, an exponent of 0.3 is used. Mathematically, the process can be written

$$S \sim E^{0.25}$$

$$\sim 10^{0.25(L/10)} \qquad \text{(above 55 dB)} \qquad (11)$$

$$S_t = S \left\{1 + F(N-1)\right\}$$

$$\sim SxFxN \qquad \text{for large N} \qquad (12)$$

$$LL_Z = 33.3 \log_{10} S_t + \text{constant}$$

$$\sim 33.3 \log_{10} F.N.S.\, 10^{0.25(L/10)} + \text{constant}$$

$$\sim 0.835\, L + 33.3 \log_{10} F.N. + \text{constant} \qquad (13)$$

The right-hand term is the asymptotic form of the perceived level increment discussed previously. However, we see that for this particular multiband noise, the calculated perceived level only increases by 8.35 dB each time the signal intensity rises 10 dB. This is true at all perceived levels above about 70 dB. The variation below that level is shown by a curve in figure 11 which has been estimated from Zwicker's loudness charts. The curve, in fact, curves downwards and crosses the unit slope line at 40 phons.

It is interesting that Žwicker's curve is very similar to Steven's function in the region of the "mid-level bulge" which is also illustrated in figure 11. The Stevens' curve is generated by both the Mk VI and Mk VII procedures; in one case by deviations in the equal loudness contour spacings, and in the other by variations in the F-factor. Note that the coincidence of the two curves at the 100 phon level is fairly realistic, but that the scale of the abscissa is based on an arbitrary reference level which differs from scale to scale.

As far as aircraft noise is concerned, we shall see that the most important feature of figure 11 is the difference at high levels between the Zwicker curve which maintains a slope of 0.835 and the remainder, which all exhibit a unit slope.

The preceding discussion has covered some of the similarities and differences between the three basic perceived level computation procedures which have formed the basis for a multitude of subsequent variations. At this stage it seems reasonable to conclude that although the basic algebraic techniques in the three approaches are very different, the net results show far more similarities than differences; particularly, if we confine our attentions to the types of levels and spectra of significance to the aircraft noise problem. In fact, the main differences between the three major techniques and their subsequent variations lie in the different frequency weighting functions.

With this background we shall proceed to examine the developments of specific methods for the evaluation of aircraft noise.

2.2 The Development of Methods for Evaluating Aircraft Flyover Noise

For the past dozen years, activity in the field of subjective response to aircraft noise has been intense. It is practically impossible, for example, to count the number of related publications which have appeared in that time. In order to maintain some continuity here, attention is confined to those studies which have contributed to, or contain significant comment upon, the development of widely used noise rating techniques. Again, it has not been possible to include every published study, and it is acknowledged that there may be significant omissions. Those that are included are treated in chronological order.

In 1959 Kryter (ref. 25) conducted an experimental study to determine how "noisy" the then new commercial jet aircraft were going to sound to people on the ground in comparison with existing propeller driven aircraft. He performed these tests with up to 100 subjects, who listened to recorded aircraft sounds played through a loudspeaker system in a large conference room. Both the method of adjustment and the method of paired comparison were used to obtain subjective ratings of the relative noisiness of 6 jet aircraft recordings and 2 piston engined aircraft recordings. A variety of rating scales were evaluated for their ability to accurately predict the judged differences. These included Stevens' loudness level, LL_S, the new Perceived Noise Level, PNL, which was based upon "equal annoyance contours", but was otherwise similar to LL_S, and four weighted sound pressure level scales. In this study, Kryter introduced the notion that noisiness and loudness are different quantities. He otherwise made the point that the different scales should only be expected to predict differences which could be attributed to spectral shape and level since these were the only factors accounted for in their specification.

The results indicated that PNL performed better than the other scales (with the exception of one of the SPL scales) since it yielded the smallest mean error between predicted and judged differences. However, it was only marginally better than the weighted SPL scales although LL_S performed particularly badly. It is important to note that in this study, the sample size was rather small, and that the scales were evaluated purely upon their ability to estimate differences between jet and piston-propeller aircraft sounds.

In 1961 Little (ref. 25), as a result of tests performed by the Boeing Company, concluded that the PNL procedure did not adequately account for the presence of intense pure tones in an otherwise broadband sound. Such a spectral combination is typical of jet engine noise which includes both compressor (fan) and exhaust components. Little experimented with as many as 150 subjects in a "large demonstration room," and synthetic sounds, including both broadband and narrow band random noise, with and without tones. He also performed some verification tests in the field using real aircraft flyover sounds as test stimuli. As a result of the study, he recommended a correction term for the PNL procedure, which has since become known as the "Little tone correction". This is shown in figure 12 and consists of an increment to be added to the normally calculated PNL as a function of the amplitude of the spike as measured with a 1/24-octave band filter.

The basic PNL procedure was revised in 1963 to include new equal noise contour data obtained by Kryter and Pearsons (ref. 15). The new curves, obtained from an experiment performed in a large classroom involving more than 200 subjects, mainly differed from the previous ones at the highest frequencies above 1000 Hz. To evaluate the new contours, a validation test was performed using nine different sounds. These sounds, which each had a duration of four seconds, were presented to a group of subjects via loudspeakers in a "semi-diffuse laboratory room." They included seven artificial sounds comprising various bands of noise, one diesel engine sound, and one recording of a jet aircraft landing. The method of adjustment was used to obtain equally noisy levels for the nine sounds and a number of scales were evaluated for their ability to predict these levels relative to that of a standard octave band (600-1200 Hz). These were LL_S, PNL, L, L_A, L_B, and L_C.

As part of the same study, Kryter and Pearsons also investigated some effects of spectral content and duration. Using octave bands of noises with and without pure tones present, they confirmed Little's finding that the tones were responsible for an increase of perceived level. However, their data, which is included in figure 12 for comparison, showed little quantitative agreement, and they concluded that further research was required. On the subject of signal duration, an experiment was performed using steady sounds of varying duration in the range 1.5 to 12 seconds measured between 10 dB-down points. They discovered that independent of signal rise and and decay times (which were varied between 1/2 and 4 seconds), for constant perceived noise level the signal level decreased very consistently by 4.5 dB each time the duration was doubled. The signal levels were fairly high, however (around 100 dB), and it was conjectured that the 4.5 dB figure might decrease with level.

Robinson (ref. 26) in 1964 analyzed the results of three separate experiments (by Copeland, et. al. (ref. 27), Robinson and Bowsher (ref. 28) and Kryter (ref. 23) to compare the relative merits of PNL (both the 1959 and 1963 versions), LL_Z, LL_S (Mk VI), L_A, and a revised version of Stevens' method which incorporated an upward masking effect similar to that of Zwicker. For a total of thirteen aircraft sounds he found Zwicker's method most consistently estimated the judged differences in perceived level, followed by LL_S (revised), PNL ('59), PNL ('63), LL_S, and L_A, in that order.

Kryter and Pearsons (ref. 29) published in 1965 the results of their further research into the effects of tones. They performed two tests with both earphone and loudspeaker sound presentation in an anechoic chamber. Twenty-one subjects used the method of paired comparison to determine the equally noisy levels of a variety of octave bands, with and without tones in the frequency range 500-6300 Hz. They found that the influence of the tone increased with frequency up to 4000 Hz and as a result recommended a tone correction procedure based on the addition of some increment to the measured SPL of the band containing the tone. This increment was specified as a function of frequency and tone-to-noise ratio on the relevant band. This could be derived in a variety of ways depending on whether the tone level could be measured independently of the noise or not. In the latter case, a tone is "identified" if the level in any particular octave, 1/3-octave or 1/10-octave band exceeds the mean level of adjacent bands by more than 0, 1 or 3 dB, respectively. The corresponding correction for 1/3-octave bands is illustrated in figure 13.

Pearsons (ref. 30) continued the earlier studies of duration, extending the total range to periods of 64 seconds. He determined that the effect of duration on perceived noisiness is a continuously varying function of level, best approximated by three straight line segments. In equivalent terms, these correspond to an increase of 6 dB per doubling in the duration range 1.5 to 4 seconds, 3.5 dB per doubling between 4 and 16 seconds, and 2 dB per doubling for durations in excess of 16 seconds. A parallel study of the effects of background noise on perceived noise level was inconclusive.

In 1966 Little and Mabry (ref. 31) reported on a series of studies directed specifically at the effects of duration on subjective response to aircraft noise. They used random noise and aircraft sounds with durations between 1 and 16 seconds, and two groups of 47 subjects. They requested one group to pay special attention to the duration of the sounds, while duration was not mentioned to the second group. They concluded that although increased duration did increase annoyance, the magnitude of the increase depended upon the test method and, particularly, was greater when the subjects were told to take duraation into account.

In early 1967, Pearsons (ref. 32) described a study of subjective response to helicopter noise. The experiments were conducted in an anechoic chamber using the method of paired comparison. Twenty-one subjects compared eight recorded helicopter sounds with two reference sounds comprising a recorded jet flyover and a wideband noise intended to simulate jet noise. As in previous studies, a number of rating scales were evaluated by testing their ability to accurately rate the judged differences between the various sounds. The scales were L, L_A, L_N, and PNL (1963). Apparently for the first time, duration and pure tone allowances were also added to the calculated PNL values in accordance with references 29 and 30. Two duration corrections were used based on the 10 dB-down and 20 dB-down durations of the N-weighted sound pressure level time histories. The study revealed that non-corrected PNL was the most consistent, although L_A and L_N were "only

slightly less accurate." Neither duration nor tone corrections improved PNL. No attempts were made to add tone or duration corrections to the other scales.

In a 1967 experiment, Pearsons and Horonjeff (ref. 33) used category scaling methods to evaluate the L, L_A, LL_S, L_N and PNL scales. Two separate tests were performed; one in an anechoic chamber using 12 recorded flyover sounds of various aircraft, and the other in the field using real flyover stimuli. In both tests the subjects were asked to rate individual sounds on an absolute category scale, i.e. very quiet, quiet, moderate, noisy, very noisy. Linear regression methods were used to correlate measured and calculated perceived levels. It was found that PNL, LL_S, and L_N were equivalent in their consistency, whereas L_A was slightly worse, and L was poor.

Also in 1967 Wells (ref. 34) reviewed research conducted by the General Electric Company and questioned the accuracy of the PNL procedure, with and without tone corrections. He proposed revisions to the equal noise contours and suggested yet another tone correction procedure, citing the evidence of subjective experiments involving 56 various synthetic sounds. He gave the revised scale the name Annoyance Level, ANL, and showed it to be a significant improvement over previous PNL procedures.

At about the same time, Mabry and Little (ref. 35) reported a further study in which they correlated perceived level estimates with "complaint potential." Seven aircraft flyover recordings were involved which required tone corrections between 3 and 11.5 dB and duration corrections between -2.5 and -6.1 dB*. They observed that ". . complaint predictions are meaningfully related to the tone correction of PNL, but not to the duration correction . . ."

In a parallel study, Little and Mabry (ref. 37) performed an experiment to validate the then current FAA recommended noise rating method (ref. 36) by comparison with sixteen other techniques. The same seven recorded sounds were used, and again the methods were rated on their performance in estimating the judged level differences between six of the sounds and one which was selected as a reference. The scales evaluated were a number of variations of LL_S and PNL involving numerous tone and duration corrections and L_A. This time, however, the standard deviation, rather than the mean of the error, was used as a criterion of performance. Little concluded that there was no significant difference between the best fifteen scales. The L_A scale was one of the two distinguished as inferior.

*The corrections were made in accordance with methods proposed by the FAA (ref. 36) for use in aircraft noise certification criteria which in turn were based on the earlier studies by Little (ref. 26) and Kryter and Pearsons (refs. 25 and 29).

In 1968 the author reported a study of subjective responses to the noise of general avaiation aircraft (ref. 9). Twenty subjects rated the recorded sounds of 35 aircraft flyovers by the method of paired comparison. The tests were conducted in a progressive wave chamber which generated essentially undirectional sound. An "absolute" judged level was determined for each sound by direct or indirect comparison with a standard reference sound*, and twenty-six rating scales were evaluated by calculating the coefficient of correlation between the calculated and judged levels. It was found that PNL with and without tone corrections, LL_Z, LL_S, and L_N performed reasonably well, but that the duration correction had no beneficial effect. Evidence provided by some aircraft flyover sounds which exhibited no Doppler frequency shift was used to suggest that this might explain the observed lack of duration effects for real aircraft noise. A further equal noise contour for free field conditions was also presented.

At the same time Hecker and Kryter released a report (ref. 38) describing another study to test the performance of various methods for rating aircraft noise. Their tests were performed with 20 subjects using loudspeakers and earphones in a "sound treated listening room." Recordings of 12 jet and 1 piston engined aircraft were rated in comparison with three reference sounds, including two octave bands of noise and one aircraft sound. Six simulated aircraft flyovers were also studied. Fifteen rating scales were tested, 12 of which were variations of the PNL procedure. The remaining three were L_A, L_N and L_C. Various conclusions were drawn regarding the relative merits of the various PNL techniques, but it is apparent that the differences were statistically small. Also, the uncorrected L_N scale performed as well as the best of the PNL scales.

In 1968 Pearsons et al. (ref. 39) attempted to shed further light onto the effect of tones present in noise spectra. They performed subjective experiments involving a wide variety of spectra with and without single and multiple, "plain" and modulated tones. They could find, however, no necessity to modify existing tone correction procedures.

Also in 1968, Hinterkeuser and Sternfeld (ref. 40) conducted a program to evaluate the synthetic sounds of six different types of future V/STOL aircraft. Eighty-two subjects used the paired comparison to rate 12 sounds heard in a semi-diffuse classroom. Six of the sounds corresponded to cruise flyover conditions, and six to longer duration terminal operations. The PNL procedure, with and without tone and duration corrections, was compared with L_A and L_C. With the possible exception of the C-scale, little difference between any of the scales was apparent.

In August 1968 and again in April 1969, Wells (ref. 41) published a description of some further possible revisions to the PNL procedure concerned with the shape of the noy curves and the form of the tone correction. Data based on 120 judged levels of various

*An octave band of "pink" noise centered at 1000 Hz.

synthetic sounds were used to demonstrate the apparent superiority of the new ANL procedure over various forms of PNL.

Also in April 1969, Young and Peterson (ref. 42) published the results of a detailed analysis of published data (refs. 23, 27 and 28) pertaining to the evaluation of aircraft noise. They concentrated upon the standard deviations of the errors between calculated and judged perceived levels to compare PNL, LL_S, L_A, L_B, L_C and L_N. They also included a new weighted sound pressure level based on a convenient electronic weighting circuit and termed LC_{28}. They could find no statistically significant difference between any of the scales except between L_B, L_C and the remainder. However, no duration or tone corrected scales were included.

In 1969, Pearsons and Bennett (ref. 43) performed yet another evaluation of various weighting scales in three experiments using: (a) 30 synthetic sounds of unusual spectral combinations, (b) twenty sounds with extreme temporal variations (e.g. durations up to 100 seconds), and (c) twenty recorded aircraft flyovers. In each case, a similar sound was used as a reference and the scales were tested by analyzing the means, standard deviations and ranges of the errors between calculated and measured perceived levels.

Nineteen different scales were evaluated, including fourteen different PNL variants, L_N, L_A (all with and without duration corrections) and L. It was concluded that the PNL procedure recently adopted by the FAA (ref. 44) for noise certification purposes was the most accurate measure for unusual sounds since it generally produced the lowest errors, however measured. It performed particularly well for the second test, involving sounds with very short to very long durations. Inspection of the detailed results, however, shows the differences between many of the scales are statistically small. This is particularly true for the aircraft noise experiment where ten other scales were at least as good, if not better in some cases, than the selected best performer.

In November 1969, Wells (ref. 45) with Pearsons (ref. 46), examined the subjective effects of sounds containing multiple pure tones. Two independent studies were made of "comb spectra", i.e. broadband noise with a large number of pure components to determine how well the various rating measures could accommodate these rather unusual sounds. The tests were performed using the method of adjustment in anechoic chambers with 20 and 30 subjects. Both investigations indicated that a newly revised version of Wells' ANL performed better than PNL with or without tone and duration corrections. However, the improvement noted by Pearsons was much smaller than that determined by Wells.

Parry and Stevens (ref. 51) at the same time published a review and analysis of previous findings regarding the applicability of the duration correction. It was their opinion that the need for this correction in aircraft noisiness ratings had not been established and that the use of such corrections should be discontinued.

However, six months later, in April 1970, a paper by Young (ref. 48) described a similar review involving 117 data points. It was concluded that despite considerable variation from test to test, a duration allowance equivalent to 3 dB per time doubling was a good average of all the data.

In August 1970, Kryter, Johnson and Young (ref. 49) described an experiment to measure the judged noisiness of a variety of fixed winged aircraft and helicopters during real flyovers. For their tests, a large number of subjects were seated out of doors or inside houses in the vicinity of the aircraft flight paths. A variety of rating scales were evaluated for their ability to predict elicited judgments of the relative perceived levels of different "reference" and "comparison" aircraft. Ninety-six subjects participated in the tests, which involved 378 separate aircraft flyovers in 189 pairs. This enabled equal perceived levels to be established for 18 different aircraft. The rating scales studied included L_A, L_B and L_C and three different forms of L_N (denoted L_D), PNL with and without a variety of tone and duration corrections, and one including a revised summation procedure. An "onset duration" correction was also tested which accounted for the rate of increase of the observed sound level which has many times been suggested as a factor of possible importance. The different forms of L_D, distinguished by the subscripts 1, 2 and 3 were based upon the original 40-noy contour, a revised contour designed to account for a "critical band effect" at low frequencies and the RC_{28} profile recommended by Young and Peterson (ref. 42). The revised PNL, abbreviated "PNdB-M" was modified to combine 1/3-octave bands so that the results approximated critical bandwidths. The study indicated that of all these measures (38 were examined), the revised PNL, with a duration correction, was the best performer. However, it was suggested that L_D(with the revised low frequency curve - D_2), with a duration correction, might be as good if not better. Also, it was admitted that ". . . because of the high correlations of the physical units themselves . . . it is something of a problem to choose a meaningful way to select those units which might be considered significantly the better predictors of subjective judgement."

In a further study of the relative merits of these 38 measurement scales, Kryter (ref. 8) combined the data from 17 previous experiments to provide a data set comprising judged perceived levels of 143 aircraft flyover sounds. Each judged level was expressed relative to that of some reference sound and the various scales were again evaluated in terms of their ability to predict the judged differences by computing the mean and the standard deviations of the errors. However, because of the same difficulties of discriminating between the various rating scales referred to above, Kryter devised a technique which essentially combined the average and standard deviation values, assuming that the two are equally important. This would indeed be true if a constant reference sound were common to all basic judgements. However, since this was not the case in this combined analysis, the mean error is a misleading statistic which is biassed by the experiments involving the greater number of data points. It is for this reason that scales with very

low standard deviations appear near the botton of the rank listing in Kryter's Table V. Disregarding his results based on the composite error, it appears that the duration corrected L_D scales proved substantially more consistent than the PNL scales, and even duration corrected L_A performed almost as well as the better PNL scales for this large set of aircraft sounds.

Summary

In order to provide a more convenient summary of this rather confusing picture of research relating to the evaluation of aircraft flyover noise, Table I has been prepared which lists the more significant facts pertaining to the various studies. These include: (i) whether the study was primarily experimental or an analysis of previous data; (ii) the type of sounds studied, aircraft (real or recorded) synthesized aircraft, or other recorded or synthetic sounds such as tones, bands of noise, etc.; (iii) what type of reference sound was used; (iv) the test method--paired comparison (PC), method of adjustment (IA), or category scaling (CS); (v) the listening conditions--semi-reverberant (SR), anechoic (A), or earphones (E); (vi) the number of subjects taking part in the experiments; (vii) the dynamic range of the sounds (in terms of peak OASPL, where possible); (viii) the effective duration of the sounds in seconds (very approximate due to a variety ot different measurement methods); (ix) the parameter used to evaluate the various rating scales--$\bar{z}$ indicates some form of mean error, s corresponds to a measure of variability, normally the standard deviation of the error, R_C is a correlation coefficient, and (x) the basic rating scales evaluated (the symbol o is peak or max*, x indicates the use of a duration correction). Again, this table is not as complete as it might be, but most of the studies responsible for the present trends in aircraft noise evaluation in the United States are included. Some facts are omitted, either because they are not relevant or because they could not be gleaned from published literature. In particular, various scales studied have been listed under a general abbreviation. For example, L_N indicates SPL scales using a weighting function based on an inverse of a noy contour. Several contours have in fact been utilized. Stevens' loudness scale, LL_S, exists in at least two commonly used forms. Most particularly, the scales labeled PNL and PNL_t (tone-corrected PNL) have been used in a wide variety of forms, through the use of different equal noisiness contours, different tone corrections, different duration corrections, and most recently, a slightly different summation procedure.

*Peak levels are the maximum instantaneous levels measured during a flyover. Max level is an alternative sometimes used which is based on the maximum levels occurring in any band regardless of when they occurred.

TABLE I. REVIEW OF PREVIOUS RESEARCH

Reference	STUDY	Expt. or Review	No. of Sounds			Ref. Sound			*	**	
			Aircraft	Synth A/C	Other	A/C	O.B.	Other	Test Method	Listening Conditions	Number of Subjects
23	Kryter 1959	E	14			x			IA/PC	SR	36
25	Little 1961	E			x		x	x	PC	SR	65-150
15	Kryter & Pearsons 1963 (i)	E	1		x			x	PC	E	14
15	Kryter & Pearsons 1963 (ii)	E	13		8		x			SR	250
26	Robinson 1964	R				x					
29	Kryter & Pearsons 1965	E			x		x		PC	E,A	21
30	Pearsons 1966	E			x			x	PC	A	18
31	Little & Mabry 1966	E			x			x	PC/CS	E	18
32	Pearsons 1967	E	8			x			PC	A	21
33	Pearsons & Horonjeff 1967 (i)	E	13	2	8				CS	A	20
33	Pearsons & Horonjeff 1967 (ii)	E	136						CS	O	42
34	Wells 1967	R/E			56			x			
37	Little & Mabry 1967	E	7			x					
9	Ollerhead 1968	E	20	15			x		PC	A	20
38	Hecker & Kryter 1968	E	13		6	x	x		PC	E,SR	20
39	Pearsons et al. 1969	E			x		x	x	IA/PC	A	40
40	Hinterkeuser & Sternfeld 1968	E		6			x		PC	SR	82
40	Wells 1968	R/E			120			x			
45	Wells 1969	R			119						
42	Young & Peterson 1969	R	52								
43	Pearsons & Bennett 1969	E	20	40		x		x	PC	A	20
46	Pearsons & Wells 1969	E				x			IA	A	20-30
48	Young 1970	R		117							
49	Kryter et al. 1970	E	13						PC	O/I	96
8	Kryter 1970	R	143								
-	Present Study	E	120				x		PC		25-32

* IA – Method of Individual Adjustment
PC – Method of Paired Comparison
CS – Method of Constant Stimulus

** SR – Semi-reverberant
E – Earphones
A – Anechoic
O – Outdoors
I – Indoors

TABLE I. CONTINUED

Reference	Dynamic Range (dB)	Duration (Seconds)	Evaluation Criteria *	Rating Methods Evaluated **										
				L	L_A	L_B	L_C	L_N	LL_S	LL_Z	PNL	PNL_t	ANL	Other
23	85-95				o		o	o	o		o			SIL, ELR
25	95-100													
15	90-110	1.5-12												
15	80-92	4	$\bar{z}$	o	o	o	o		o	o	o			
26	89-107		$\bar{z}$		o				o	o	o			Revised LL_S
30	86-88	4-64												
31	76-100	1-16												
32	82-95	2.6-16	$\bar{z}$, s	o	o			o			ox			
33	94-103			o	o			o	o			ox		
33				o	o			o						
34			$\bar{z}$, s								o	o	o	
37	95	7-17	s		o				o		ox	ox		Various
9	78-88	1-16	R_c	o	ox	o	o	ox	o	o	ox	ox		L_{NN}
38	89-103	7-25	$\bar{z}$		o		o	o			ox	ox		
39														
40					o						ox	ox		
40			$\bar{z}$, s								o	o	o	
45			$\bar{z}$, s	o	o	o	o	o	o	o	o	ox	o	
42			s	o	o	o	o	o		o				RC_{28}
43	64-91	1-100	z, s	o	ox			ox			ox	ox		
46		4									o	o	o	
48		0.5-75												
49	97-116	5-27	z		ox	o	o	ox	o		ox	ox		RC_{28}
8			[$\bar{z}$+ s]		ox	o	o	ox	o		ox	ox	ox	RC_{28}
-	84-115	1-27	$\bar{z}$, $\bar{s}$, R_c	ox	ox	ox		ox	ox	ox	ox	ox		L_{NN}

* $\bar{z}$ - Measure of error associated with central value
s - Measure of scatter of error
R_c - Correlation coefficient

** o - Peak level scale; x - duration allowance included.

Several clear facts emerge from this review of previous work, as follows:

1. Since the earliest experiments involving aircraft noise, there has been a preponderence of emphasis upon the PNL scales. It has been extensively studied, revised, extended and varied in the minutest detail in attempts to improve correlation with experimental observation. Such attention has not been devoted to the other basic scales. For example, the duration correction had not been applied by any investigator except the author to any of the sound pressure level scales until as late as 1969. Duration allowances have apparently never been used with the two basic loudness level procedures in connection with aircraft noise. Indeed, Zwicker's scale has hardly been used at all.

2. Most scales and their refinements have been developed on the basis of studies involving steady state or highly controllable laboratory generated (synthetic) sounds, often with extreme spectral and temporal features. Few studies indicate any real, statistically significant differences between a multitude of variations upon the basic scales when they are used to rate real or recorded aircraft flyover sounds.

3. Individual experiments have usually produced only a few data points which, because of the extremely variable nature of the problem, have represented inadequate statistical samples. The potential sources of error, even under the most highly controlled test conditions, are numerous and many experiments, particularly those performed in "semi-diffuse" test rooms or out of doors, are prone to particularly large noise measurement errors. In this regard, a common practice of defining individual sound pressure levels to the nearest one-tenth decibel, implying the same order of accuracy, is perhaps misleading. It is doubtful that, in general, test levels can be measured to an accuracy of better than 0.5 dB at a specific, calibrated point in space. Over a distributed area, particularly in outdoor conditions, very much greater errors can be anticipated.

Although several of the reported studies have been completed since the initiation of this project, it was the rather confused background described in this section that led to its inception. There seemed to be a clear need for a thorough study to determine whether, in fact, there has been any real progress in the search for better ways of measuring aircraft. Accordingly, it was decided to obtain a very large set of consistent subjective measurements of aircraft flyover sounds so that the relative merits of the basic noise rating scales could be assessed with a high degree of confidence.

3.0 EXPERIMENTAL APPROACH

3.1 Program Objectives

At the outset of this investigation it was planned to perform two basic experiments. The first was addressed at the effects of aircraft motion, to which, it has previously been conjectured (ref. 9), subjective response might be linked. The reasoning behind this hypothesis was based on many conflicting observations of the apparent effects of signal duration. As discussed in section 2.2 some investigators found a need for a duration allowance in the noise rating procedures, others found the opposite. An examination of published data revealed a clear trend amongst these findings. Experiments involving synthetic, laboratory generated sounds strongly support the need for a duration allowance, whereas experiments involving a total of more than 250 aircraft flyover sounds reveals no positive, beneficial effect of a duration correction. Thus, the evidence was that reaction to synthetic sounds increases with duration. For "natural" aircraft flyover sounds it does not. A possible explanation for this was suggested (ref. 9) by the subjective comparison of the sounds of real flyovers and recordings of aircraft which had been tailored to exhibit flyover-type level variations, but excluded the frequency changes associated with the sound of a moving aircraft. A distinct duration effect in the case of the latter could not be identified for the natural flyovers and an almost automatic conclusion was that the differences could be attributed to the Doppler effect. It is not suggested that a frequency change, per se, necessarily has any subjective magnitude, but merely that the Doppler shift conveys information to the listener about the distance and speed of the aircraft. These, in turn, might be related to the phenomenon of "perceptual constancy" which has a well-known visual counterpart and has been suggested on several occasions as a possible auditory factor of importance by Robinson (e.g. ref. 50), who refers to a "projection hypothesis." Simply, it means that a listener may tend to rate a sound with respect to what he expects to hear, rather than to some absolute reference. Thus, as an aircraft approaches, the listener expects the level to increase, and judges it to be less noisy by a corresponding amount. In a similar way, an observer might be inclined to rate equally the different sound levels of two identical aircraft flying at different altitudes.

Another reaction which might be associated with distance and motion could be called, for want of a better term, "fear reaction." This is related to such responses as fear, excitement or exhilaration produced by aircraft proximity, speed and possibly other parameters. Comments like "hair tingling" have been elicited from test subjects who have listened to recordings of nearby aircraft in flight, and , of course, the threat of accident is a known factor in real community situations.

Thus, to determine whether source motion itself might contribute to perceived level, an experiment was designed around a set of computer generated sounds. These sounds were tones which were precisely varied in level and frequency according to the Doppler

equations. Both apparent source distance and velocity were varied and relative subjective levels were obtained by the method of adjustment. The results of this experiment did indicate some trends to suggest that perceived level may vary with distance and velocity; however, the trends were too small for any definite conclusions to be drawn. For this reason, the remainder of the report is instead devoted to the second and major experiment.

The objective of this experiment was to obtain the largest possible set of experimentally measured perceived levels of aircraft flyover sounds and to correlate the judgements with the levels calculated by various noise rating procedures. In this way it would hopefully be possible to: (a) make a realistic comparison of these various procedures; (b) find out how well they performed in an absolute sense; and (c) make recommendations for further improvements which might improve their validity for rating the sounds of aircraft.

3.2 Experiment Design

The main requirement of the experiment was to generate a large number of subjective measurements of aircraft noise levels. The tests were to be performed in the laboratory using recorded sounds, and it was obviously desirable to use as wide a selection of sounds as possible. The sounds should certainly show large variations in spectral characteristics, sound pressure level, and duration in order to provide the best possible chance of discriminating between the effects of these variables. In addition, it was intended to search for any effects which might be associated with extra-auditory variables such as aircraft type, distance, and speed so that the sounds from a wide variety of aircraft recorded under a wide variety of conditions should be included. Each sound was to be judged on a meaningful scale so that the noise rating scales could be tested for absolute as well as relative accuracy.

Since the program was proceeding in parallel with a study of subjective response to STOL aircraft, being performed by Wyle Laboratories for the Federal Aviation Administration (ref. 10), the opportunity was taken to incorporate data from that investigation into the same analysis. In fact, the experimental techniques for both studies were designed to be identical so that the two sets of results would be completely compatible. The techniques are described fully in reference 10 and although for the sake of completeness the main features of the combined experiment are described here, the reader is referred to the above report for more complete information on the experimental details.

3.2.1 <u>Selection of sound recordings.</u> - A total of 120 sounds were selected from various sources including an in-house tape library, a number of recordings made specially for the two studies, and from tapes made available by various government and industrial agencies. The sounds were divided roughly equally into the four major categories: jets (turbojet and turbofan), propeller turbine powered aircraft, piston engined aircraft, and helicopters.

The sounds included outdoor recordings of flyovers, take-offs and landings with the microphone located at various positions with respect to the flight path so that the sounds comprised a wide assortment of those which might be heard on or around a mixed-traffic airport. Table II is a complete listing of the 120 sounds, which includes known or estimated data describing the aircraft type and classification, the flight mode, the slant distance between the aircraft and the microphone at its nearest point of approach, and the peak sound pressure level at the microphone location. Omissions from the table indicate that the corresponding variables are not known. Histograms showing distributions of the major variables are presented in figure 14.

3.2.2 Paired comparison procedure. - In order to rank the sounds of an absolute scale of judged perceived level, each was compared either directly or indirectly with a "standard reference" sound consisting of an octave band of "pink" noise* centered at a frequency of 1000 Hz. The measured perceived level of each could then be expressed in terms of the sound pressure level of the standard reference when it was judged to have an equal perceived magnitude. This is, of course, closely related to the basic definitions of many of the scales for calculating perceived level, particularly PNL. In fact, the judged level, obtained in this way, only differs from a measured PNL in the amount the bandwidth of the standard reference differs from an ideal octave due to the filter skirts and a finite signal-to-noise ratio.

The basic test method employed in the subjective experiments was a paired comparison technique which had been developed and evaluated in earlier studies (references 9 and 57). In common with those and other studies, the subjects were asked to evaluate the sounds with respect to noisiness where the adjective "noisy" was alternatively described as "unwanted," "objectionable," or "disturbing."

In a single paired comparison, the aircraft sound in question ("comparison"), was compared to a reference sound by asking the subjects to rate one with respect to the other during ten repetitions of the pair. In five of these, the reference (variable level) sound appeared first; in the remainder, the comparison (fixed level) was first. In each of these two sets, the reference was played at five different levels, at increments of 5 dB over a range within which the "equally noisy" level was estimated to lie. The two orders of presentation were used so that a natural subjective bias towards the sound of a pair could be eliminated by averaging. The pairs were randomly mixed with pairs associated with other comparisons so that the subjects could not recognize any regular presentation pattern.

The two sounds of each pair were separated by a one-second interval, and successive pairs by six seconds. The subjects were asked to rate the relative noisiness of the two sounds on a scale of ± 5 arbitrary units, using positive values if they considered the second sound

---

* Pink noise has a uniform spectrum level as measured by a constant percentage bandwidth analysis.

TABLE II. AIRCRAFT SOUND RECORDINGS USED.

Log No.	Aircraft Type	Propulsion Category*	Gross Weight (Lb.)	Engines No.	Engines Max.H.P. Ea. (Thrust-Lb.)	Flight Mode **	Minimum Distance (Feet)†	Estd. Speed (mph)	Measd. Peak SPL (dB)
4	Piper Cherokee	P	3,400	1	260	T	300	90	94
6	Piper Aztec	P	5,200	2	250	T	100	135	90
8	Learjet Model 23	TJ	12,500	2	(2,850)	T	1,560	190	100
9	Douglas DC9	TF	98,000	2	(14,000)	T	1,170	200	99
10	Martin 404	P	40,000	2	2,500	T	1,000	120	101
12	Douglas DC-68	P	106,000	4	2,500	T	720	130	106
13	Boeing B-727	TF	160,000	3	(14,000)	T	1,190	200	107
15	Douglas DC-8	TF	315,000	4	(18,000)	T	2,170	200	101
16	Twin Bonanza	P	7,300	2	275	T	400	100	97
17	Boeing B-707	TF	257,000	4	(14,500)	T	1,760	200	102
18	Queen Air B-88	P	8,800	2	380	T	400	100	102
19	Piper Apache	P	3,800	2	160	T	500	90	94
20	Beech Bonanza	P	3,400	1	285	T	400	90	98
21	Boeing B-707	TF	257,000	4	(14,500)	T	1,600	200	99
22	Twin Cessna	P	5,200	2	260	T	400	130	96
23	Douglas DC-3	P	26,000	2	1,200	F	(1,000)	165	92
25	Cessna Model 172	P	2,300	1	145	T	500	90	89
26	Douglas C-124	P	75,000	4	2,500	F	(750)	200	97
27	Piper Apache	P	3,800	2	160	T	400	90	93
28	Cessna Model 172	P	2,300	1	145	T	400	90	90

* TJ= Turbojet; TF= Turbofan; TP= Turboprop; P= Pisfon; H= Helicopter. ** T= Takeoff; F= Flyover; L= Landing.

† Estimated (± 5%) from photograph where available. Figures in parentheses denote visual estimates by test personnel. Omitted entries indicate unavailable data. SPL's are peak values recorded in field.

TABLE II. (CONTINUED)

Log No.	Aircraft Type	Propulsion Category	Gross Weight (Lb)	Engines No.	Engines Max.H.P.Ea. (Thrust-Lb.)	Flight Mode	Minimum Distance (Feet)	Estd. Speed (mph)	Measd. Peak SPL (dB)
30	Douglas DC-9	TF	98,000	2	(14,000)	L	220	150	97
31	Boeing B-727	TF	160,000	3	(14,000)	L	210	150	102
33	BN-2A Islander	P	5,700	2	260	T	(200)	75	109
34	BN-2A Islander	P	5,700	2	260	F	(200)	150	98
35	BN-2A Islander	P	5,700	2	260	T	(200)	75	108
36	BN-2A Islander	P	5,700	2	260	F	(200)	157	98
38	Dorn. Skyservant	P	8,050	2	380	T	(200)	75	113
39	Dorn. Skyservant	P	8,050	2	380	T	(200)	75	112
41	F-H Heliporter	TP	6,100	1	575	T	(300)	75	102
42	F-H Heliporter	TP	6,100	1	575	F	(400)	157	90
43	F-H Heliporter	TP	6,100	1	575	T	(300)	75	102
44	F-H Heliporter	TP	6,100	1	575	T	(300)	75	102
46	Helio Stallion	TP	5,100	1	680	T	(350)	50	112
47	Helio Stallion	TP	5,100	1	680	F	(350)	50	99
48	Helio Stallion	TP	5,100	1	680	T	(350)	75	109
49	Helio Stallion	TP	5,100	1	680	F	(350)	185	103
50	Helio Courier	P	3,400	1	295	F	(350)	38	105
51	Helio Courier	P	3,400	1	295	F	(350)	165	101
53	Boeing B-727	TF	160,000	3	(14,000)	T	(200)	200	113
54	BAC One-Eleven	TF	87,000	2	(11,400)	T	(200)	200	115
55	Boeing B-727	TF	160,000	3	(14,000)	T	(200)	200	111
57	McD Douglas 188	TP	58,422	4	1,480	F	1,900	250	91
59	McD Douglas 188	TP	58,422	4	1,480	F	807	250	99
60	McD Douglas 188	TP	58,422	4	1,480	F	331	250	102
61	McD Douglas 188	TP	58,422	4	1,480	F	320	250	104

TABLE II. (CONTINUED).

Log No.	Aircraft Type	Propulsion Category	Gross Weight (Lb.)	Engines No.	Engines Max.H.P.Ea. (Thrust-Lb.)	Flight Mode	Minimum Distance (Feet)	Estd. Speed (mph)	Measd. Peak SPL (dB)
63	McD Douglas 188	TP	58,422	4	1,480	T	500	75	98
65	LTV XC-142	TP	37,474	4	3,080	L	(500)	120	119
68	Grumman Mohawk	TP	19,230	2	1,100	F	1,000	300	86
69	Grumman Mohawk	TP	19,230	2	1,100	F	300	300	92
71	Grumman Mohawk	TP	19,230	2	1,100	F	1,000	300	94
74	Martin 404	P	40,000	2	2,500	L	800	100	104
76	Queen Air A-65	P	7,700	2	340	L	850	100	101
79	Martin 404	P	40,000	2	2,500	L	700	100	101
81	Lockheed Electra	TP	116,000	4	4,050	L	306	140	99
84	Bell Cobra AH-IG	H	9,500	1	1,400	F	213	50	100
85	Bell Cobra AH-IG	H	9,500	1	1,400	F	120	50	107
87	Short Skyvan III	TP	12,449	2	715	L	(500)	100	107
88	Boeing B-707	TF	257,000	4	(14,500)	L	200	150	103
89	DHC-5 Buffalo	TP	41,000	2	3,060	F	750	80	90
90	DHC-5 Buffalo	TP	41,000	2	3,060	F	750	150	95
58	McD Douglas 188	TP	58,422	4	1,480	F	814		98
70	Grumman Mohawk	TP	19,230	2	1,100	F	50		108
72	Grumman Mohawk	TP	19,230	2	1,100	F	300		103
77	Martin 404	P	40,000	2	2,500	L	650		102
78	Lockheed Electra	TP	116,000	4	4,050	L	620		100
200	Vertol CH-47B	H	33,000	2	2,850	F	100	46	109
202	Bell UH-IB	H	9,500	1	1,100	F	1,000	92	88
206	Vertol CH-47B	H	33,000	2	2,850	F	1,100	115	92
207	Vertol CH-47B	H	33,000	2	2,850	F	250	115	103
208	Kaman HH-43	H	6,100	1	770	F	200	80	84

TABLE II. (CONTINUED).

Log No.	Aircraft Type	Propulsion Category	Gross Weight (Lb.)	Engines No.	Engines Max.H.P.Ea. (Thrust-Lb.)	Flight Mode	Minimum Distance (feet)	Estd. Speed (mph)	Measd. Peak SPL (dB)
211	Kaman HH-43	H	6,100	1	770	L		107	107
212	Kaman HH-43	H	6,100	1	770	L		92	92
213	Kaman HH-43	H	6,100	1	770	T		86	86
214	Bell Cobra AH-IG	H	9,500	1	1,400	F	85	60	109
216	Bell UH-ID	H	9,500	1	1,100	F	153	50	107
217	Gyrodyne QH-50	H	2,300	1	300	F	125	35	
220	Gyrodyne QH-50	H	2,300	1	300	F	1,000	46	
221	Hughes OH-6	H	2,400	1	215	F	125	57	
222	Hughes OH-6	H	2,400	1	215	F	1,000	69	
223	Hughes OH-6	H	2,400	1	215	F	500	115	
225	Sikorsky CH-3E	H	18,500	2	1,400	F	125	69	
226	Sikorsky CH-3E	H	18,500	2	1,400	F	1,000	69	
227	Sikorsky CH-3E	H	18,500	2	1,400	F	500	115	
229	Bell UH-IF	H	9,000	1	1,039	F	125	69	
230	Bell UH-IF	H	9,000	1	1,039	F	1,000	69	
232	Bell UH-IF	H	9,000	1	1,039	F	500	115	
234	Kaman HH-43	H	6,100	1	770	F	125	69	
235	Kaman HH-43	H	6,100	1	770	Hover			
236	Bell UH-IF	H	9,000	1	1,039	F			
250	Douglas DC-9	TF	98,000	2	(14,000)	T	660	200	108
251	Douglas DC-9	TF	98,000	2	(14,000)	T	875	200	105
252	Douglas DC-9	TF	98,000	2	(14,000)	T	930	200	105
253	Douglas DC-8	TF	315,000	4	(18,000)	T	780	200	111
254	Douglas DC-9	TF	98,000	2	(14,000)	T	570	200	110
255	Douglas DC-8	TF	315,000	4	(18,000)	T	580	200	105

TABLE II. (CONCLUDED).

Log No.	Aircraft Type	Propulsion Category	Gross Weight	Engines No.	Engines Max.H.P.Ea. (Thrust-Lb.)	Flight Mode	Minimum Distance (feet)	Estd. Speed (mph)	Measd. PEAK SPL (dB)
256	Boeing B-727	TF	160,000	3	(14,000)	T	630	200	107
257	Boeing B-727	TF	160,000	3	(14,000)	T	1,300	200	107
258	Grum. Gulfstream	TP	35,000	2	2,190	T	930	150	101
259	Douglas DC-9	TF	98,000	2	(14,000)	T	1,700	200	104
260	Douglas DC-6B	P	106,000	4	2,500	T	1,700	130	102
261	Boeing B-727	TF	160,000	3	(14,000)	T	1,850	200	98
262	Douglas DC-9	TF	98,000	2	(14,000)	T	1,800	200	99
263	BAC One-Eleven	TF	87,000	2	(11,400)	T	1,750	200	100
264	H-Siddley HS-125	TJ	20,500	2	(3,360)	T	1,800	200	99
265	Douglas C-124	P	75,000	4	2,500	T	1,700		105
266	Boeing B-727	TF	160,000	3	(14,000)	T	1,750	200	99
267	Convair 580	TP	55,000	2	3,750	L	200	150	101
268	Douglas DC-9	TF	98,000	2	(14,000)	L	200	150	104
280	Lockheed Electra	TP	116,000	4	4,050	L	400		
281	Lockheed Electra	TP	116,000	4	4,050	T	600		
282	DHC-5 Buffalo	TP	41,000	2	3,060	F	750		
283	DHC-5 Buffalo	TP	41,000	2	3,060	F	750		
284	Douglas DC-9	TF	98,000	2	(14,000)	T	(200)		
285	Boeing B-727	TF	160,000	3	(14,000)	T	(200)		
286	Douglas DC-9	TF	98,000	2	(14,000)	T	(200)		
287	Boeing B-727	TF	160,000	3	(14,000)	T	(200)		
288	BAC One-Eleven	TF	87,000	2	(11,400)	T	(200)		
289	BAC One-Eleven	TF	87,000	2	(11,400)	T	(200)		
290	Boeing B-727	TF	160,000	3	(14,000)	T	(200)		
291	Boeing B-727	TF	160,000	3	(14,000)	T	(200)		

to be more objectionable than the first, and negative numbers to indicate that they considered the second less objectionable than the first. Each subject recorded his scores on an IBM Portapunch card for subsequent computer analysis. The written instructions given to the subjects are shown on the following page. These were verbally reiterated by the Test Director at the time of the tests, and every attempt was made to ensure that the instructions were understood by all participants.

3.2.3 Sound pair arrangements. - Although each sound could be compared directly with the standard octave band reference sound, it has generally been found undesirable to do so because (1) it is difficult to comparatively judge aircraft flyby sounds with respect to random noises, and (2) the constant repetition of a particular sound causes a rapid increase in the subjects' error rate. Consequently, use was made of a number of intermediate reference sounds so that most aircraft noises were related to the standard reference by way of a two-stage comparison. Although this causes some accumulation of errors, these are not as serious as those referred to above.

Thus, a system of paired comparisons was devised comprising three levels of reference, as shown in figure 15. The simulated jet noise (resembling that of a jet engine ground runup), served as a "half-way point" between the Level 3 and 4 references and the standard. The 1/3-octave band spectra of the Level 1 and Level 2 references are shown in figure 16. The Level 3 aircraft noise reference was selected to exhibit as short a duration as possible to minimize the difficulty of comparison with the four-second reference sounds.

The particular comparison arrangement was selected so that the perceived levels of the Level 2 and Level 3 references, with respect to that of the standard, could be carefully established by replicated measurements and by closing a number of 1-2, 2-3, 3-1 comparison triangles. The latter allowed a convenient check on the consistency of the subjective judgements.

Altogether, 150 sound pair comparisons were arranged. Each comparison involved 10 pairs of sounds so that the total experiment required 1500 sound pairs. These were divided into 60 tests, each test containing 25 pairs of sounds and lasting about 15 minutes, which has been found to be an acceptable duration for a single sitting. The method by which the sixty test tape recordings were prepared is described in Appendix A.

3.2.4 Test procedure. - The sounds were played to either 25 or 32 subjects in Wyle Laboratories' progressive wave acoustic chamber, which has been described in detail in references 9 and 10. Figure 17 shows a cutaway view of this facility in its configuration for the present experiments. Four or five subjects are seated facing an array of loudspeakers. Two separate systems were used, which included five 60-watt low-frequency units to generate frequencies below 500 Hz and a single multi-cellular horn for the higher frequency range. A set of 12 foot-long fiberglass wedges was installed behind the subjects to absorb the essentially unidirectional sound waves which are uniform across the facility test section to within ± 3 dB at all frequencies. The sound system, which is

INSTRUCTIONS - PLEASE READ VERY CAREFULLY

The purpose of these tests is to find out how people react to the sound of certain aircraft. More specifically, we are trying to measure how NOISY these aircraft are. In using the word NOISY we are referring to sound that could be called alternatively UNWANTED, OBJECTIONABLE, or DISTURBING. Whatever its interpretation, NOISINESS cannot be measured easily with instruments because it is basically a subjective value. Consequently, it is only through tests of this kind that we really evaluate the disturbance caused by aircraft noise.

You will record your test results directly onto punched cards, which in turn will be computer processed. Please ensure therefore that you punch these cards correctly; any error which you may make can readily be corrected by the Test Director if you draw it to his attention at the end of a test. Such errors may, however, be difficult to correct later. Before the test begins please check your assigned subject and seat number in the appropriate spaces on the left-hand side of the punched card.

Each test will comprise twenty-five pairs of sounds, with a short break between each sound of a pair and a longer break between pairs. A lighted indicator in front of you will identify which sound pair is currently being played, and will also show whether it is the first or the second sound of the pair. For each of the twenty-five pairs, there is a corresponding numbered column on your score card to record your response.

FOR EACH PAIR OF SOUNDS, YOUR JOB IS TO DECIDE WHETHER THE SECOND SOUND IS MORE NOISY OR LESS NOISY THAN THE FIRST, AND TO PUNCH AN APPROPRIATE SCORE IN THE CORRECT COLUMN ON YOUR CARD

If you think the second sound is

- MORE NOISY, punch a POSITIVE (+) NUMBER,
- LESS NOISY, punch a NEGATIVE (-) NUMBER

Use a number which is appropriate to the difference between the two sounds, so that 1 represents a very slight difference, 2 a greater one and so on. Note that any number up to 5 can be used.

For example, if you consider the second sound to be slightly more noisy than the first, you will probably punch +1. If, on the other hand, you believe the second sound to be very much less noisy than the first, you may perhaps punch -5.

If you feel that the two sounds are equally disturbing punch the number zero (0).

PLEASE REMEMBER THAT YOU ARE JUDGING THE SECOND SOUND RELATIVE TO THE FIRST. YOU MAY THINK THAT BOTH SOUNDS ARE NOISY, OR THAT NEITHER IS PARTICULARLY NOISY. NEVERTHELESS YOU MUST DECIDE WHETHER THE SECOND SOUND DISTURBS YOU MORE THAN THE FIRST .

In making your judgments please listen to both sounds completely and try to take account of all the effects the sounds might have upon you if heard in your home many times during the day and night. Please record an answer for every pair of sounds, even though you feel you may be guessing. There is no right or wrong answer. All that is required is your own personal opinion.

Should you make an error simply repunch the correct score IN THE SAME COLUMN and point out the correct result to the Test Director when the test has finished.

IT IS MOST IMPORTANT THAT YOU FULLY UNDERSTAND THESE INSTRUCTIONS. IF YOU ARE AT ALL UNCERTAIN ABOUT WHAT YOU ARE TO DO, PLEASE ASK THE TEST DIRECTOR TO EXPLAIN AT ANY TIME.

The value of the experiment would be reduced by any panel member's failure to fully comprehend the scoring system, or being confused by the card punching procedure. Please FEEL FREE to question the Test Director.

capable of generating sound pressure levels in excess of 114 dB in the frequency range 20-5000 Hz and above 118 dB between 25 and 4000 Hz is described in Appendix A.

In order to maintain a uniform frequency response throughout the overall system, a spectrum shaper was included in the replay circuit. This was used to adjust the frequency response to give a flat spectrum at the center of the test section when pink noise was inserted at the signal source.

The thirty-two panel members were chosen to provide a reasonable distribution of age, occupation, and sex, the selection being made on the basis of audiometric measurements and an aptitude test. This was designed to test ability to make consistent judgements in a paired comparison test. For this purpose, a special tape was made which contained a triangulated comparison test. Three individual aircraft sounds were used to form the three pairs 1-2, 2-3, 3-1. The test was administered to sixty audiometrically screened applicants simultaneously. The final panel selected comprised 20 females and 12 males with a median age of 26 years. The 25 members of the smaller panel were selected from the larger group. Individual tests were performed with four or five persons at a time, and took a total of eight weeks to complete.

3.2.5 Data reduction. - The subjects recorded a total of 42,750 individual scores which were processed by a series of computer programs described in reference 10.

For each sound pair comparison and each level of the reference sound, each subject's two scores, J_1 and J_2, for "forward" and "reverse" orders of presentation, were summed, remembering that the sequence reversal effectively changes the sign of the score. Thus,

$$\bar{J} = J_1 - J_2$$

where $\bar{J}$ = mean score in arbitrary units

J_1 = score (from +5 to -5) assigned by the subject to that level with the reference presented first

J_2 = score when the reference was presented second

Most subjects have a tendency to overemphasize the level of the second pair, and this summing procedure compensates for this "order effect." The bias toward the second sound was, in fact, found to be, on average, equivalent to about 1 dB. Histograms of the mean scores for all subjects were generated by the computer, and examples are shown in figure 18. Two complete sets of five are included, corresponding to increasing levels of the reference sound from left to right of the figure.

Computer programs were also written to fit curves through the median points of the five histograms in each set in order to obtain a "zero crossing," at which relative level it is assumed the two sounds would be judged equally noisy. Curves were also fitted to the 20th and 80th percentile points to give some measure of the scatter of the data about this intercept. This analysis is described in reference 10 and examples are shown in figure 19. However, subsequent to the completion of the STOL noise study, a reexamination of the subjective data showed that the use of the median scores and automatic procedures sometimes led to rather crude curve fits, particularly when the intercept occurred a long way from the origin. Accordingly, the entire set of data was analyzed for the present purposes by hand fitting curves to the computed means rather than medians. This approach produced slightly different, but hopefully improved, results to those obtained previously (ref. 10). The zero-crossing intercepts were used, in conjunction with sound pressure level plots recorded during each of the test sessions through a monitoring microphone located at the center of the facility test section, to derive an equivalent "equally noisy" level of the standard reference sound for each of the aircraft sounds.

3.3 Acoustic Data Analysis - Objective Measurements

Each of the aircraft noise recordings was analyzed to yield various objective measures of perceived level. This was done in two stages, using specially developed computer programs: (i) processing of the acoustic signals into frequency-time matrices of 1/3-octave band sound pressure levels; and, (ii) conversion of these matrices into time varying and integrated estimates of perceived level. The two steps will be described separately.

3.3.1 One-third octave band analysis. - For the purposes of calculating the various objective measures described in Section 3.3.2, it was necessary to reduce each flyover sound to a matrix $L_{1/3}$ (n, k), where

$L_{1/3}$ = "instantaneous" 1/3-octave band sound pressure level in dB

n = 1/3-octave band number

k = time increment counter corresponding to successive 1/3-second intervals.

The 120 aircraft sounds were recorded in sequence onto master analysis tapes, using identical procedures to those used for making the test tapes themselves (see Appendix A). These tapes were played into a bank of 1/3-octave band filters (B&K Model 123 covering the 24 center frequencies between 50 and 10 kHz. (Frequencies outside this range were, in fact, filtered from the signals played to the subjects). The individual

outputs from each of the filters were then fed to two multiplexer/A-D converter units for high speed acquisition by an XDS Sigma V computer system. The total digital conversion rate was approximately 95,000 samples per second, the minimum required to compute accurate mean square levels for all 24 channels. The signal voltages were squared and averaged and stored at 1/2-second intervals in "real-time" for later dB conversion and calibration by a second computer program. This converted each mean square voltage to an accurate 1/3-octave band sound pressure level, using an appropriate calibration factor for each band. Each band level was then specified in dB relative to the same arbitrary reference for each frequency band.

3.3.2 Objective perceived level computations. - The second stage of the analysis was to compute, from each of the 1/3-octave band levels spectra, a number of perceived level ratings. Although a very large number of alternative scales were contenders for inclusion in the analysis, it was necessary for practical purposes to restrict these to a relatively small number. Consequently, attention was confined to a small number of variations of the three major procedures for perceived level estimation. These are: (i) the weighted sound pressure level scale; (ii) Stevens' loudness level calculation; and, (iii) Zwicker's loudness level scale. Although thirty-six different variations were computed, including twelve basic scales, each with two forms of "duration correction," only eighteen have been retained for discussion in this report. These include nine basic scales, with and without duration corrections. The remainder are omitted for reasons which will be discussed in section 4. Each of those retained and the reasons for their inclusion will be described in turn.

As discussed in section 2.0, all the scales involve some form of frequency band summation procedure of the form

$$PL(k) = \sum_{n=1}^{24} f_n\left[L_{1/3}(n, k)\right] \qquad (14)$$

where n = band number (1 to 13 between 50 and 10,000 Hz)

k = time increment counter

$L_{1/3}$ = 1/3-octave band sound pressure level

$PL(k)$ = the calculated perceived level at the k-th time instant

and f_n is some function of frequency and level which varies from scale to scale. For each scale and each sound under study, two values of perceived level were retained for correlation with the subjective data. These are PL, the largest value of PL(k) occurring during the flyover and referred to as simply the peak level, and EPL, a time integrated or

"duration corrected" value of PL(k) which is referred to as an "effective" level. This is calculated according to the summation equation

$$EPL = 10 \log_{10} \frac{1}{T} \sum_{k=K_1}^{K_2} 10^{PL(k)/10} \Delta t \tag{15}$$

where Δt, the time increment between samples = 0.5 sec; K_1 and K_2 correspond to the time intervals when PL(k) first and last exceeds a level which is 10 dB below the peak level PL (these are automatically replaced by 0 and/or K the first and last values if the 10 db-down points do not occur in a particular record), and T is a reference time of 10 seconds.

Each of the scales used to compute PL(k), together with their abbreviations and the reasons for their inclusion, are described below in turn.

Overall Sound Pressure Level, L

This is the simplest sound level scale in use since it consists simply of the sound pressure level on linear or non-frequency weighted scale in units of dB re: $2 \times 10^{-5} N/M^2$.

$$L(k) = 10 \log_{10} \left\{ \sum_{n=1}^{24} \text{antilog}_{10} \frac{L_{1/3}(n,k)}{10} \right\} \tag{16}$$

The overall sound pressure level scale is known to be a poor predictor of perceived level. It is concluded in the present analysis to act as a base-line against which to compare the remainder of the methods.

A and B Weighted Sound Pressure Levels, L_A, L_B

These are the frequency weighted sound pressure level scales which are available on standard sound level meters (ref. 12). In this analysis they are estimated according to the formula

$$L_{A,B} = 10 \log_{10} \left\{ \sum_{n=1}^{24} \text{antilog}_{10} \frac{L_{1/3}(n,k) - W_{A,B}(n)}{10} \right\} \tag{17}$$

Values of the weighting functions W_A and W_B are given with other functions in Table III. The C-scale has been omitted since it only involves slight deviations from a zero weighting function at extreme ends of the frequency range and in all applications L_C has proved practically identical to the linear scale L.

N-weighted Sound Pressure Level, L_N. (Sometimes referred to as the D-scale.)

This uses a weighting function which is the inverse of the 40-noy contour (e.g. reference 23). Similar to the B-weighting at low frequencies, it involves a characteristic "hump," enhancing the importance of frequencies above 1000 Hz. It has frequently been found to be as good a predictor of perceived level as the more complex PNL procedure.

NN-weighted Sound Pressure Level, L_{NN}

This scale, evaluated previously in reference 9, incorporates a weighting function based on an equal noisiness contour measured in the same progressive wave facility in which the present tests were conducted. Although this function attributes an unusually large degree of importance to frequencies in the region of 4000 Hz, its form was confirmed in a subsequent, independent experiment (ref. 51) and it did rank highly in previous comparisons with other scales for application to aircraft noise (ref. 9). Although "non-standard," it has been included out of curiosity to see if the same performance could be repeated.

Stevens' Loudness Level, Mk VI, LL_S

This method, described in reference 22, was discussed in detail in section 2. By definition, the calculated loudness level of a complex sound in phons should numerically equal the sound pressure level in dB of a 1000 Hz tone which is judged to be equally loud. The computer subroutine used in this analysis incorporated mathematical representations of the "loudness index" charts presented in that reference, together with conversion equations given therein. For a detailed description of the procedure, the reader is referred to Stevens' paper. It should be noted that this procedure has now been supplemented by a much refined Mk VII version (ref. 2).

TABLE III

TABLE OF WEIGHTING FUNCTIONS

Frequency Hz	Attenuation, dB			
	L_A	L_B	L_N	L_{NN}
50	30	12	12	11
63	26	10	11	10
80	22	7	9	9
100	19	6	7	9
125	16	5	6	8
160	13	3	5	7
200	11	2	3	6
250	8	2	2	5
315	6	1	1	4
400	5	0	0	3
500	3	0	0	3
630	2	0	0	2
800	1	0	0	1
1000	0	0	0	0
1250	-1	0	-2	-1
1600	-1	0	-6	-2
2000	-1	0	-8	-4
2500	-1	0	-10	-8
3150	-1	0	-11	-15
4000	-1	0	-11	-20
5000	-1	1	-10	-20
6300	0	2	-9	-15
8000	1	3	-6	-8
10,000	3	5	-3	0

Zwicker's Loudness Level, LL_Z (Approximate Method)

Zwicker's loudness level has been computed by a subroutine developed on the basis of the "graphical evaluation charts" of reference 19. However, the method adopted is only an approximation of the procedure since the effect of masking has been ignored. This was justified on the grounds that for the majority of aircraft noise spectra, the contributions to loudness of the spread of masking is probably small. This assumption was certainly confirmed by a small number of comparisons of computed and graphically obtained levels where the errors were less than 0.5 phon. However, significant discrepancies could occur in the presence of spectral "spikes" and it must be admitted that this approximate approach may not be as accurate as the rigorous procedure. Also, the parameters programmed are only as accurate as readings which could be made from published documents.

Perceived Noise Level, PNL

The particular version of perceived noise level utilized was originally described by Kryter and Pearsons (ref. 15). Like LL_S, it relates to a subjective definition that the perceived noise level of a sound, in PNdB, is numerically equal to the sound pressure level of the "equally noisy" octave band of noise centered at 1000 Hz. The equations programmed are based on mathematical formulations of the noy tables developed by Pinker (ref. 52) and specified by the FAA in reference 44. The table is accordingly programmed as a two-segment equation of the form

$$N(n, k) = \text{antilog}_{10} \quad M \left\{ L_{1/3}(n, k) - L_0 \right\} \tag{18}$$

where M and L_0 are functions of both the band number N and the level $L_{1/3}(n, k)$ as tabulated in reference 44.

The noy values are combined by the relationship

$$N_t(k) = N_{max}(k) + 0.15 \left[\left\{ \sum_{n=1}^{24} N(n, k) \right\} - N_{max}(k) \right] \tag{19}$$

This is then transformed to perceived noise level in PNdB as

$$PNL(k) = 40 + 33.3 \log_{10} N_t(k) \tag{20}$$

Tone-Corrected Perceived Noise Level PNL_t

This scale corrects the basic PNL for the presence of pure tones in the spectrum. Of several alternatives, the method used is that adopted by the FAA, which in turn was based upon an original recommendation by Little (ref. 26). Previous studies have indicated that this technique is perhaps preferable to suggested alternatives. "Tones" in the spectrum are identified by initially smoothing the spectrum to a broadband "background" level and comparing smoothed and unsmoothed band levels. If the latter exceeds the former by 3 dB or more, a tone correction increment is specified as a function of frequency (n) and the height of the "tone projection." The largest correction derived in this manner is then added to PNL(k) calculated as indicated above. A detailed description of these tone correction procedures may be found in reference 44. The integrated version of this scale, $EPNL_t$, is that specified by the FAA for aircraft noise certification measurements.

The 1/3-octave band sound pressure level arrays output by the spectrum analysis program were referred for convenience to some arbitrary datum. The second program, before computing the peak and effective perceived levels, corrected the arrays to the standard reference of 2×10^{-5} N/M^2 by reference to the known measured peaks overall levels for each signal. This was done simply by computing the overall level time history, L(k), from the uncalibrated data and comparing the peak level, L, with the measured peak level to obtain a decibel error. This constant was then added or subtracted from each value of $L_{1/3}(n, k)$ before computing the remaining perceived levels. An example of the four-part computer output is shown in Table IV. The third sheet gives the time histories of the various calculated levels and the asterisks denote: (1) levels which are within 10 dB of the peak; (2) the extremities of the 10 dB-down range; and, (3) the peak level in each case.

Although it is not possible to reproduce the detailed results of all the acoustic analyses, plots of the overall sound pressure level histories and the 1/3-octave band sound pressure level spectra corresponding to the peak overall level are presented for all 120 sounds in Appendix B.

The complete set of perceived levels, both judged and calculated, is listed for each sound in Table V. It is important to note that the peak levels, L, listed in Table V, are those measured during playback to the test subjects, and thus differ somewhat from the original recorded levels presented in Table II. Since the differences are normally small, however, correct perceived levels for the original sounds, should they be required, can be accurately obtained by adjusting those in Table V for the differences in peak level, L.

TABLE IV. CALCULATION OF OBJECTIVE NOISE LEVELS.

LOG. NO. 59   -185 FOURTH FLYOVER, 807 FEET ALTITUDE

NAERO = 22
NDT = 42
EXPSPL = 100

L(N,K) MATRIX OF THIRD OCTAVE BAND LEVELS.
(CORRECTED FOR CALIBRATION FACTORS ONLY)

N = T	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	67	73	73	63	68	62	65	72	73	72	66	65	62	62	59	58	57	53	50	50	52	54	51	48
2	66	73	73	66	67	62	67	71	70	71	64	62	59	63	60	59	56	52	49	51	52	54	51	48
3	69	78	75	62	67	64	71	75	77	70	67	65	61	62	57	59	56	52	51	55	52	53	50	47
4	71	74	68	65	64	67	69	74	74	70	73	67	64	63	60	62	58	53	51	51	52	54	51	48
5	65	77	73	70	71	71	73	75	75	69	76	70	69	66	63	63	59	56	54	54	55	54	51	48
6	68	74	75	67	67	73	74	79	78	69	76	67	70	65	65	66	62	56	55	53	53	54	51	48
7	71	71	75	63	70	75	75	80	74	74	77	68	68	66	64	67	61	56	53	55	54	54	51	48
8	70	74	74	65	71	77	76	76	75	79	77	69	66	67	67	67	63	57	56	54	52	54	51	48
9	77	77	74	64	72	77	73	79	75	82	77	72	69	69	67	68	64	61	59	56	53	55	52	49
10	75	76	76	67	78	79	78	77	80	82	76	71	72	67	70	69	66	63	63	59	55	54	51	48
11	69	76	75	68	80	81	82	78	85	81	83	75	74	71	72	69	68	64	67	61	56	55	52	49
12	71	70	73	77	86	79	76	84	87	88	84	78	76	74	74	71	70	68	69	61	57	55	52	49
13	73	86	77	82	90	80	82	87	89	90	83	80	79	75	74	72	67	67	67	59	57	56	53	50
14	80	92	78	84	94	79	87	89	93	87	82	77	75	74	76	72	69	71	67	62	58	57	54	51
15	88	90	75	93	97	85	90	89	91	81	83	77	75	76	75	71	70	74	66	62	58	58	55	52
16	81	75	75	97	85	83	90	86	93	81	84	80	77	79	76	69	73	71	65	61	58	58	55	52
17	83	73	81	91	80	83	88	88	89	80	82	80	78	80	75	71	74	68	63	60	57	57	54	51
18	90	73	80	79	79	72	86	87	87	84	85	80	81	81	77	69	71	65	61	58	56	56	53	50
19	83	72	83	80	82	80	88	90	81	85	81	77	84	81	75	68	67	61	59	56	55	56	53	50
20	84	72	84	78	82	84	80	85	81	88	79	78	83	80	72	67	68	59	58	56	55	55	52	49
21	80	74	75	75	84	83	76	82	81	85	79	76	80	76	68	65	64	59	55	54	53	55	52	49
22	76	72	67	73	78	81	79	75	79	72	78	76	77	76	68	63	63	57	53	52	52	53	50	47
23	74	75	65	67	72	84	87	75	77	75	79	74	76	72	66	62	60	56	53	52	52	54	51	48
24	69	65	69	67	74	81	83	76	74	79	74	76	76	71	66	59	61	55	53	54	52	54	51	48
25	73	73	76	61	76	79	84	78	69	74	72	72	76	69	62	59	57	54	51	52	52	54	51	48
26	67	72	75	60	71	78	86	79	68	74	69	72	70	68	58	59	57	52	52	52	53	54	51	48
27	68	73	74	60	71	76	81	79	68	70	69	71	71	69	58	57	57	53	52	52	53	55	52	49
28	70	72	73	60	68	75	82	79	73	68	71	69	70	65	56	55	57	53	51	51	52	54	51	48
29	74	75	75	59	66	73	79	79	71	66	70	66	69	64	56	55	56	52	50	51	52	54	51	48
30	70	75	75	66	73	71	83	75	71	65	69	65	68	63	56	55	57	52	50	51	53	55	52	49
31	63	70	76	64	57	67	78	76	71	63	67	61	64	61	55	54	55	51	51	55	53	54	51	48
32	66	73	76	64	73	69	78	73	70	69	62	67	61	61	55	56	57	51	52	50	53	54	51	48
33	72	69	75	63	69	66	76	75	71	68	61	65	60	58	54	52	56	50	50	51	52	54	51	48
34	65	69	70	65	64	62	78	70	69	67	59	64	59	58	53	53	55	51	50	51	52	53	50	47
35	61	67	73	67	75	59	75	70	68	67	59	60	58	58	53	52	54	50	50	51	52	55	52	49
36	65	67	75	62	60	61	70	74	69	68	59	64	60	60	54	53	59	50	51	54	52	55	52	49
37	67	71	73	64	71	57	67	70	69	64	61	60	62	57	52	53	56	51	50	51	53	56	53	50
38	65	73	70	63	73	58	66	71	67	62	60	56	58	55	53	53	56	51	50	51	51	54	51	48
39	63	69	72	64	65	59	61	68	65	64	57	62	59	57	53	53	55	50	50	51	53	54	51	48
40	74	74	70	69	74	66	65	65	64	61	61	57	60	58	54	55	55	52	53	52	54	56	53	50

TABLE IV. (CONTINUED).

THIRD-OCTAVE LEVEL CORRECTION FACTOR = -1

L(N,K) MATRIX OF THIRD-OCTAVE BAND LEVELS

(WITH OVERALL LEVEL CORRECTION FACTOR)

N = / T	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	66	72	74	62	67	61	64	71	72	71	65	64	61	61	58	57	56	52	49	49	51	53	50	47
2	65	72	72	65	66	61	66	70	69	70	63	61	58	62	59	58	55	51	48	50	51	53	50	47
3	68	77	74	61	66	63	70	74	76	69	66	64	60	61	56	58	55	51	50	54	51	52	49	46
4	70	73	67	64	63	66	67	73	73	69	72	66	63	62	59	61	57	52	50	50	51	53	50	47
5	65	76	72	69	70	70	72	74	74	68	75	69	68	65	62	62	58	55	53	53	54	53	50	47
6	67	73	77	66	60	72	73	78	77	68	75	66	69	64	64	65	61	55	54	52	52	53	50	47
7	70	70	74	62	69	74	74	79	73	73	76	67	67	65	63	66	60	55	52	54	53	53	50	47
8	69	73	73	64	70	76	75	75	74	78	76	68	65	66	66	66	62	56	55	53	51	53	50	47
9	76	76	77	63	71	76	77	78	74	81	76	71	68	68	66	67	63	60	58	55	52	54	51	48
10	74	75	77	65	77	78	77	76	79	81	77	70	71	66	69	68	65	62	62	57	54	53	50	47
11	68	75	77	67	79	80	81	77	84	80	82	74	73	70	71	68	67	63	66	60	55	54	51	48
12	70	69	72	76	85	78	75	83	86	87	83	77	75	73	73	70	69	67	68	60	56	54	51	48
13	72	85	76	81	89	79	81	86	88	89	82	79	77	74	73	71	66	66	66	58	56	55	52	49
14	79	91	77	83	93	78	86	88	92	86	81	76	74	73	75	71	68	70	66	61	57	56	53	50
15	87	89	74	92	96	84	89	88	90	80	82	76	74	75	74	70	69	73	65	61	57	57	54	51
16	80	74	77	96	64	82	89	85	92	80	83	79	76	78	75	68	72	70	64	60	57	57	54	51
17	82	72	80	90	79	82	87	87	88	79	81	79	77	79	74	70	73	67	62	59	56	56	53	50
18	89	72	79	76	78	71	85	86	86	83	84	79	80	80	76	68	70	64	60	57	55	55	52	49
19	87	71	79	79	81	79	87	89	80	84	80	76	83	80	74	67	66	60	58	55	54	55	52	49
20	83	71	83	77	81	83	79	84	80	87	78	77	82	79	71	66	67	58	57	55	54	54	51	48
21	79	73	74	74	83	82	75	81	80	64	78	75	79	75	67	64	63	58	54	53	52	54	51	48
22	75	71	66	72	77	80	78	74	78	71	77	75	76	75	67	62	62	56	52	51	51	52	49	46
23	73	74	64	66	71	83	86	74	76	74	78	73	75	71	65	61	59	55	52	51	51	53	50	47
24	68	64	68	66	73	80	82	75	73	78	73	75	75	70	65	58	60	54	52	53	51	53	50	47
25	72	72	75	60	75	78	83	77	68	73	71	71	75	68	61	58	56	53	50	51	51	53	50	47
26	66	71	74	59	70	77	85	76	67	73	68	71	69	67	57	58	56	51	51	51	52	53	50	47
27	67	72	73	59	70	75	80	78	67	69	68	70	70	68	57	56	56	52	51	51	52	54	51	48
28	69	71	72	59	67	74	81	78	72	67	70	68	69	64	55	54	56	52	50	50	51	53	50	47
29	73	74	74	58	65	72	78	78	70	65	69	65	68	63	55	54	55	51	49	50	51	53	50	47
30	69	74	74	65	60	70	82	74	70	64	68	64	67	62	55	54	56	51	49	50	52	54	51	48
31	62	69	75	63	56	66	77	75	70	62	66	60	63	60	54	53	54	50	50	54	52	53	50	47
32	65	72	75	63	72	68	77	72	69	68	61	66	60	60	54	55	56	50	51	49	52	53	50	47
33	71	68	74	62	68	65	75	74	70	67	60	64	59	57	53	51	55	49	49	50	51	53	50	47
34	64	68	69	64	63	61	77	69	68	66	58	63	58	57	52	52	54	50	49	50	51	52	49	46
35	60	66	72	66	61	58	74	69	67	66	56	59	57	57	52	51	53	49	49	50	51	54	51	48
36	64	56	74	61	59	60	69	73	68	67	58	63	59	59	53	52	58	49	50	53	51	54	51	48
37	66	69	72	63	70	56	66	69	68	63	60	59	61	56	51	52	55	50	49	50	52	55	52	49
38	64	72	72	62	72	57	65	70	66	61	59	55	57	54	52	52	55	50	49	50	50	53	50	47
39	52	68	71	63	64	58	60	67	64	63	56	61	58	56	52	52	54	49	49	50	52	53	50	47
40	73	73	69	68	73	65	65	64	63	60	60	56	59	57	53	54	54	51	52	51	53	55	52	49

TABLE IV. (CONTINUED).

AIRCRAFT NUMBER 22 LOG NUMBER 59 MD-186 FOURTH FLYOVER, 807 FEET ALTITUDE

MAX NOISE LEVEL	100	103	106	100	102	90	96	100	96	95	96	94
AT TIME INTERVAL	15	14	15	14	14	14	15	15	15	14	15	15
INTEGR. DURATION CORRECT.	-6.5	-4.6	-5.0	-4.1	-3.1	-3.4	-5.0	-6.6	-5.0	-4.6	-5.1	-4.9
APPROX. DURATION CORRECT.	-4.8	-3.3	-4.0	-2.2	-1.3	-2.7	-3.3	-4.8	-3.3	-3.0	-3.6	-3.0
CORRECT. MAXLEVEL (INTEGR)	93.5	98.4	101.0	95.9	98.9	86.6	91.0	93.4	91.0	90.4	90.9	89.1
CORRECT. MAXLEVEL (APPROX)	95.2	99.7	102.0	97.8	100.7	87.3	92.7	95.2	92.7	92.0	92.4	91.0

TABLE IV. (CONCLUDED).

AIRCRAFT NUMBER 22 LOG NUMBER 59 MD-186 FOURTH FLYOVER, 807 FEET ALTITUDE

TIME INTERVAL NUMBER	OASPL	PNL(K)	PNLT(K)	LL(S)	LL(Z)	SPL(A)	SPL(B)	SPL(C)	SPL(D)	SPL(J1)	SPL(J2)	SPL(NN)
1	[illegible]	85	86	85	89	73	77	80	78	77	78	78
2	79	85	85	85	88	72	76	79	77	76	77	78
3	83	87	86	87	90	75	79	83	80	79	79	80
4	81	87	88	86	90	75	79	81	79	79	79	79
5	83	89	89	88	92 **	77	81	83	81	81	81	81
6	85	90	92	89	93 *	78	82	85	83	82	83	82
7	85	91	92	89	93 *	79	83	85	83	82	83	82
8	85	91	92	90 **	94 *	79	83	85	84	83	83	82
9	87	94 **	97 **	92 *	95 *	81 **	85	87	86 **	85 **	85	84 **
10	88	94 *	95	92 *	96 *	82 *	86 **	88	87 *	86 *	86 **	86 *
11	[illegible] **	97 *	99 *	94 *	98 *	85 *	89 *	90 **	89 *	89 *	89 *	88 *
12	93 *	99 *	101 *	97 *	100 *	87 *	92 *	93 *	92 *	91 *	92 *	91 *
13	95 *	101 *	101 *	98 *	101 *	89 *	94 *	96 *	94 *	93 *	94 *	92 *
14	96 *	103 ***	103 *	100 ***	102 ***	90 ***	95 *	98 *	95 *	95 ***	95 *	93 *
15	[illegible] ***	103 *	106 ***	100 *	102 *	90 *	96 ***	100 ***	96 ***	95 *	96 ***	94 ***
16	99 *	103 *	106 *	100 *	102 *	90 *	96 *	99 *	96 *	95 *	96 *	93 *
17	98 *	101 *	104 *	98 *	101 *	89 *	93 *	96 *	93 *	92 *	93 *	91 *
18	95 *	99 *	104 *	97 *	101 *	90 *	92 *	94 *	92 *	92 *	92 *	91 *
19	95 *	[illegible] *	105 *	97 *	100 *	90 *	92 *	95 *	92 *	92 *	92 *	90 *
20	93 *	99 *	103 *	97 *	99 *	88 *	91 *	93 *	91 *	91 *	91 *	89 *
21	91 **	97 *	100 **	94 *	97 *	85 *	89 *	91 **	89 *	88 *	89 *	87 *
22	87	92	93	91 *	95 *	82 *	85	87	85	85 *	85	84 *
23	89	94 **	95	92 *	95 *	82 *	88 *	89	87 **	86 *	87 **	85 *
24	87	92	93	91 *	94 *	81 *	86 **	87	85	85 **	85	84 **
25	87	92	93	90 *	93 *	80 **	85	87	84	84	84	83
26	87	92	93	90 **	92 *	79	85	87	85	84	85	83
27	85	[illegible]	91	88	92 **	77	82	85	82	81	82	81
28	85	[illegible]	91	88	91	77	83	85	82	82	82	81
29	84	[illegible]	89	87	90	77	81	84	81	80	81	80
30	85	[illegible]	90	88	90	76	82	85	82	81	82	80
31	82	87	89	86	89	73	79	82	79	78	79	80
32	82	87	85	86	89	73	79	82	79	78	79	79
33	81	[illegible]	89	85	88	73	78	81	78	77	78	78
34	[illegible]	[illegible]	87	85	87	71	77	80	77	77	77	78
35	79	[illegible]	87	84	87	70	76	79	76	75	76	77
36	79	[illegible]	83	84	88	71	76	79	77	76	77	78
37	[illegible]	[illegible]	85	83	87	70	74	78	75	75	75	77
38	79	[illegible]	85	83	86	69	74	79	75	74	75	77
39	[illegible]	81	81	82	86	68	72	76	74	73	74	77
40	[illegible]	83	84	84	87	72	74	79	75	74	75	78

TABLE V. JUDGED AND CALCULATED PERCEIVED LEVELS (dB)
(Aircraft Sounds Identified by Log Number)

LOG	JPL	L	EL	L_A	EL_A	L_B	EL_B	L_N	EL_N	L_{NN}	EL_{NN}	LL_S	ELL_S	LL_Z	ELL_Z	PNL	EPNL	PNL_t	$EPNL_t$
4	91	96	90	88	81	93	87	93	86	91	84	97	91	100	93	100	93	104	96
6	94	94	92	89	85	91	88	96	91	100	95	100	95	103	99	103	98	105	101
8	98	101	98	96	93	100	96	100	97	98	95	104	101	106	104	107	103	109	105
9	95	99	99	94	92	97	96	98	96	98	95	102	100	106	103	105	103	107	105
10	94	101	95	93	89	98	93	98	93	95	90	102	97	105	100	105	100	110	103
12	105	113	108	105	101	110	105	110	105	109	104	116	110	116	112	118	113	119	115
13	99	108	104	99	96	105	102	104	102	101	99	108	106	110	107	111	108	115	110
15	99	103	101	101	98	103	100	104	101	102	100	109	106	111	109	111	108	115	110
16	91	96	91	90	83	94	88	94	88	92	86	99	93	101	95	101	95	107	97
17	103	106	105	101	99	104	102	105	103	105	102	109	107	112	110	112	110	114	112
18	103	111	105	105	100	109	103	109	104	110	105	116	110	115	111	118	112	123	116
19	94	98	92	91	86	94	89	96	91	99	92	100	96	104	100	103	98	108	102
20	93	100	95	92	84	96	90	96	90	97	91	101	96	105	98	104	98	104	99
21	95	99	99	94	92	97	95	98	95	97	94	102	100	106	103	105	102	109	104
22	93	97	93	91	85	95	90	95	90	95	88	101	95	103	97	103	97	104	99
23	90	91	88	83	80	88	85	89	85	88	83	94	90	97	93	96	92	97	94
25	87	88	84	83	75	86	80	86	80	85	79	92	86	95	89	94	87	94	87
26	93	103	99	97	93	98	93	98	93	96	92	103	99	105	101	106	101	110	104
27	86	92	85	78	71	87	80	86	79	85	78	92	85	94	86	93	86	100	91
28	83	92	87	81	74	87	82	86	80	84	79	91	86	93	87	91	87	99	90
30	107	103	96	102	94	102	94	110	101	116	107	114	105	115	106	117	108	122	112
31	112	111	102	108	99	109	110	115	106	121	112	120	111	120	111	122	113	123	114
33	102	111	102	99	90	106	97	106	97	108	99	112	103	112	103	114	105	116	107
34	88	96	87	85	77	92	83	93	84	97	87	99	90	101	93	101	92	104	94

TABLE V. CONTINUED

LOG	JPL	L	EL	L_A	EL_A	L_B	EL_B	L_N	EL_N	L_{NN}	EL_{NN}	LL_S	ELL_S	LL_Z	ELL_Z	PNL	EPNL	PNL_t	$EPNL_t$
35	108	118	111	113	104	116	109	118	110	122	113	125	116	123	115	127	118	129	120
36	97	101	95	96	87	99	91	102	93	107	97	107	98	110	101	110	101	111	102
38	110	114	107	109	103	113	106	116	109	121	114	121	114	121	115	123	116	126	118
39	113	115	108	110	104	113	107	117	110	123	116	122	116	122	115	123	117	126	120
41	101	106	100	94	88	101	96	101	96	100	96	106	101	105	102	108	103	111	106
42	86	95	85	86	78	92	83	92	83	91	84	96	88	99	92	99	91	101	93
43	100	108	103	99	93	104	99	104	99	103	99	109	104	110	105	112	106	118	111
44	94	103	99	93	88	99	94	99	94	97	94	103	99	104	101	106	102	112	106
46	100	115	105	106	96	112	102	112	102	110	101	116	107	115	106	118	109	125	114
47	93	103	93	97	86	101	91	101	91	99	90	105	96	106	98	108	98	115	105
48	99	114	104	107	97	113	102	112	102	110	101	117	107	116	107	119	109	121	112
49	98	111	100	102	92	109	98	108	98	106	97	112	102	111	102	115	105	122	111
50	99	104	97	95	89	102	95	101	95	102	95	106	100	107	103	109	103	114	108
51	94	99	90	93	85	97	88	97	89	101	90	102	94	106	98	105	97	112	101
53	108	114	108	108	102	112	105	113	106	117	108	119	112	119	112	120	114	122	116
54	111	115	109	110	104	114	107	115	109	117	109	120	114	119	114	122	116	123	117
55	110	115	110	111	104	114	107	115	109	119	110	121	114	120	114	123	116	130	119
57	86	93	91	84	83	89	88	89	88	86	86	94	93	95	95	96	95	100	98
59	92	100	94	90	87	96	91	96	91	94	89	100	96	102	99	103	98	106	101
60	98	104	98	99	92	102	96	104	98	102	96	108	102	110	104	111	104	114	106
61	98	106	97	99	91	104	95	104	95	102	94	109	100	109	102	112	103	113	105
63	99	102	100	94	93	99	97	99	97	97	95	103	101	104	103	106	104	110	107
65	107	116	113	110	105	115	111	114	110	112	108	120	115	118	115	122	117	122	119
68	87	88	82	85	80	87	82	93	96	99	90	96	90	99	93	99	91	101	94

TABLE V. CONTINUED

LOG	JPL	L	EL	L_A	EL_A	L_B	EL_B	L_N	EL_N	L_{NN}	EL_{NN}	LL_S	ELL_S	LL_N	ELL_N	PNL	EPNL	PNL_t	$EPNL_t$
69	95	94	88	91	84	93	87	97	90	104	95	102	95	105	99	105	98	107	100
71	91	96	89	92	83	95	87	95	87	93	86	99	92	100	94	102	94	105	99
74	98	104	99	94	90	100	96	100	95	98	94	105	100	105	102	107	103	111	105
76	88	99	93	85	82	94	89	93	89	92	87	98	94	97	95	100	95	104	98
79	97	99	95	92	88	96	92	96	92	95	91	101	97	103	100	103	100	108	102
81	102	104	97	102	95	102	95	104	97	105	98	111	104	111	105	112	105	112	107
84	97	102	98	97	93	96	92	97	92	97	92	102	98	105	100	105	100	106	102
85	95	107	101	99	95	104	96	104	96	103	96	109	101	110	102	112	103	115	106
87	98	111	101	96	88	105	96	106	97	109	101	111	102	109	102	113	104	120	110
88	107	103	95	104	96	103	95	113	104	117	109	116	108	114	106	119	110	123	115
89	82	87	85	83	80	83	80	83	81	84	82	89	88	93	92	90	88	93	91
90	85	94	90	90	84	90	84	90	84	87	83	93	90	95	92	96	92	97	92
58	93	95	91	88	84	92	88	92	88	90	86	97	93	100	96	99	95	105	98
70	103	106	94	106	94	106	94	112	99	115	102	117	104	117	105	120	107	124	111
72	101	106	98	100	93	104	96	105	99	112	104	110	104	112	106	113	107	118	110
77	97	102	96	92	86	98	93	98	92	97	91	103	97	104	99	106	100	108	102
78	97	99	95	97	91	98	93	100	95	101	95	106	101	107	103	107	102	111	105
200	96	108	103	101	96	106	99	105	99	103	98	109	104	110	107	112	107	114	110
202	94	96	95	87	87	93	92	92	92	90	90	97	97	100	99	100	99	103	101
206	96	98	97	89	87	95	93	94	93	92	91	98	97	99	98	100	99	104	103
207	93	102	95	93	87	97	92	96	91	95	90	101	97	103	99	104	99	107	102
208	93	90	85	85	79	88	82	89	83	90	83	95	89	99	93	96	90	97	92
211	89	101	96	94	89	98	94	98	94	96	93	102	99	104	102	105	101	107	103
212	98	98	96	89	88	94	93	95	93	94	92	99	98	103	101	102	100	104	102

TABLE V. CONTINUED

LOG	JPL	L	EL	L_A	EL_A	L_B	EL_B	L_Z	EL_Z	L_{NN}	EL_{NN}	LL_S	ELL_S	LL_Z	ELL_Z	PNL	EPNL	PNL_t	$EPNL_t$
213	101	97	92	97	91	97	92	101	96	101	97	104	99	108	104	108	102	109	103
214	102	108	104	101	96	103	99	102	98	101	97	107	103	110	104	110	105	112	108
216	94	101	99	96	94	98	94	98	96	103	98	104	101	107	104	106	103	107	105
217	96	91	90	84	84	87	87	88	88	92	91	95	95	99	99	96	96	101	99
220	87	86	84	78	77	83	81	83	81	84	80	89	88	93	92	90	88	93	91
221	95	101	95	92	87	98	92	98	92	97	91	102	97	105	100	105	100	112	104
222	96	94	94	85	85	91	91	90	91	88	89	95	95	97	98	97	97	102	102
223	90	93	90	84	82	90	87	90	87	87	85	94	92	97	96	96	94	102	98
225	95	101	97	95	90	97	92	99	94	100	96	104	100	108	103	107	102	108	104
226	86	88	87	79	80	83	82	84	82	83	82	90	90	94	93	92	90	95	93
227	93	97	92	86	84	92	89	93	89	93	90	99	96	101	99	101	97	104	100
229	94	101	100	94	93	96	95	97	95	100	96	103	100	106	103	105	103	109	105
230	89	87	88	82	81	86	85	86	85	87	83	92	91	96	94	94	92	97	95
232	93	98	97	92	90	91	91	90	90	89	89	95	95	97	97	97	97	100	100
234	95	91	86	85	81	88	83	89	85	92	88	95	91	100	96	97	93	100	95
235	83	84	82	75	73	79	78	80	79	83	81	87	86	91	90	88	86	89	88
236	89	92	93	84	84	85	86	84	85	83	84	89	91	92	92	91	91	93	93
250	101	108	103	100	96	105	100	106	101	111	103	110	106	113	108	113	109	114	111
251	104	101	98	95	93	98	96	99	97	204	98	104	101	107	105	106	104	109	107
252	101	101	100	96	93	98	97	101	98	105	99	104	102	109	106	108	105	111	108
253	104	107	105	100	98	104	102	106	104	110	105	110	108	112	110	114	111	117	114
254	103	108	103	101	97	104	101	107	102	113	105	111	107	114	109	115	110	118	112
255	104	106	102	101	95	104	99	107	101	114	105	112	106	114	108	115	109	118	111
256	96	108	102	99	95	106	100	105	100	106	100	110	105	111	107	112	107	115	109

TABLE V. CONTINUED

LOG	JPL	L	EL	L_A	EL_A	L_B	EL_B	L_N	EL_N	L_{NN}	EL_{NN}	LL_S	ELL_S	LL_Z	ELL_Z	PNL	EPNL	PNL_t	$EPNL_t$
257	97	102	97	94	90	98	94	98	94	96	93	102	99	106	102	105	101	108	104
258	95	102	96	87	82	97	90	96	89	94	89	101	95	99	96	103	97	106	101
259	94	95	93	89	87	93	91	93	91	91	89	98	96	102	100	100	98	102	100
260	90	96	93	87	83	90	87	89	86	88	85	95	92	96	93	96	93	101	96
261	95	94	94	90	87	93	90	93	90	91	88	97	95	101	98	100	97	103	100
262	87	91	89	86	83	90	86	90	86	88	84	95	92	98	95	97	93	99	96
263	89	90	89	87	84	89	87	90	88	89	86	94	92	99	96	96	94	100	97
264	87	90	85	87	82	89	84	90	85	89	83	95	90	99	94	96	91	99	94
265	97	107	103	102	98	101	98	101	98	99	97	106	103	108	105	108	106	114	109
266	89	90	88	84	82	87	85	87	86	85	84	92	91	97	95	94	93	98	95
267	106	103	95	103	95	102	94	113	105	122	114	115	107	109	103	117	109	124	115
268	109	97	90	97	88	96	88	105	96	112	102	108	99	109	102	110	102	112	104
280	96	96	89	89	84	89	85	94	89	100	94	100	95	102	98	101	96	103	98
281	90	96	87	79	74	87	80	87	81	90	85	94	88	95	90	94	89	96	90
282	88	92	88	87	83	87	83	87	83	85	82	93	89	95	92	94	90	95	92
283	93	97	92	93	87	94	87	94	88	91	87	97	93	99	96	100	95	102	96
284	93	93	90	88	85	90	88	95	90	102	94	99	95	102	99	102	97	105	99
285	94	95	94	90	88	92	91	96	92	102	93	100	97	103	100	103	99	103	101
286	92	91	88	85	83	87	85	91	87	98	89	96	92	99	96	98	94	103	96
287	93	91	90	86	84	88	87	93	89	99	92	97	94	101	98	100	96	101	98
288	93	97	95	92	88	94	91	95	92	99	93	101	97	104	100	103	99	104	101
289	92	90	89	86	84	89	87	92	88	96	88	95	93	99	97	97	95	98	97
290	94	96	96	89	87	91	90	92	91	95	91	97	97	102	100	100	98	101	99
291	95	97	96	91	89	94	93	95	94	100	95	100	98	104	102	103	101	105	103

4.0 ANALYSIS AND DISCUSSION OF RESULTS

4.1 Correlation Statistics

Section 3 described how the perceived level of each sound was measured in two ways: (a) subjectively by analyzing the responses of a group of people exposed to the sounds, and (b) objectively by physical analysis of the acoustic signals themselves. Before proceeding to compare the two sets of values, it is necessary to establish precisely what we expect of the various rating scales.

4.1.1 Sources of error. - Ideally, the objective level should be identical to the subjective level in every case. In reality, this ideal is rarely achieved and the two will differ by an amount which varies from sound to sound and from scale to scale. These differences represent the cumulative effects of several errors:

1) The error due to inherent subjective variability.

2) The error due to the inability of the scale to accurately account for all the characteristics of the sound which are subjectively important.

3) The experimental error associated with inaccurate measurements and analyses of both the subjective and objective data.

Little can be done about errors (1) and (3) beyond taking all normal precautions to avoid errors and to ensure that the test panel is fully trained in its task. The second error is the quantity we are trying to measure and the problem at hand is to distinguish between this error and the other two which are always present.

The subjective variability (2) can be estimated in a very approximate manner from the errors associated with the zero-crossing intercepts of the paired comparison results. Examples are shown in figure 19 where the intersections of the 20th and 80th percentile curves can be seen on either side of the median intersection. An analysis of a large set of zero crossing data (reference 10) showed that the average range between the 20th and 80th percentile intercepts was approximately 9.5 dB. Assuming this to be approximately equal to two standard deviations, a standard t-test would indicate the 95 percent confidence intervals associated with the mean zero crossing to be ± 1.7 dB. That is, in many repetitions of any test, we may be confident that 95% of the intercepts would lie within the range ± 1.7 dB. This is the error associated with the group of 32 subjects. For the smaller group, the error will be a little larger. Also note that these figures correspond to the average behavior of the test panel. It may also be inferred that any individual would perform to this degree of accuracy. However, if the test repetitions were based on individual results, repeatedly picked at random from the group, the equivalent confidence interval would be ± 6.5 dB, a range which is, therefore, more representative

of the error to be expected for any individual with normal hearing and picked at random from any population.

A measure of the total experimental error which effectively includes both (1) and (3) has been obtained from the 20 comparisons of the Level 1 and Level 2 reference sounds included in the tests and described in Section 3.2.3, where the standard deviation of the distribution of results was 1.5 dB. Also, in the ten 1-2, 2-3, 3-1 triangulation checks also described in section 3.2.3, the average magnitude of the ten errors in closing the loops was 0.9 dB. Thus, it seems that subjective variability is the greatest source of experimental error whose r.m.s. value lies between 1 and 2 dB.

4.1.2 Single variable error analysis. - Accepting that the probable r.m.s. experimental error is of the order 1-2 dB, it may be assumed that deviations between the calculated and judged perceived levels which exceed this range may be attributed to the inadequacies of the perceived level scale. In any event, there is no way in which the two sources of error can be separated, and they can be analyzed only in combination.

The value of a rating scale, of course, rests with its ability to accurately and consistently estimate the perceived noise level of aircraft flyover noise. In the present context, "accuracy" could be used to describe the absolute agreement between the judged and calculated levels, whereas "consistency" might refer to the dispersion of the errors about some central value. Thus, for example, the scale which repeatedly yields calculated levels of 90 dB for a number of sounds which are all subjectively rated at 100 dB, is not at all accurate, but very consistent. If we assign the variables x and y to the calculated and judged perceived levels of a sound, as will be done through the remainder of this report, we could determine accuracy and consistency from the distribution of the error (x-y). Accuracy is related to the mean error

$$\bar{z} = \frac{\sum_{i=1}^{N} (x_i - y_i)}{N} \tag{21}$$

whereas consistency is reflected by the sample standard deviation of the error which is given by

$$s = \left[\frac{\sum_{i=1}^{N} (x_i - y_i)^2 - \left\{ \sum_{i=1}^{N} (x_i - y_i) \right\}^2 / N}{N - 1} \right] \tag{22}$$

where N is the number of samples, x_i and y_i are the objective and subjective levels associated with the i-th sample.

The practical distinction between the term "accuracy" and "consistency" rests with our ability to define the judged levels in appropriate terms and in this regard it is essential to recognize the importance of a reference point. In writing down equations (21) and (22), it has been assumed that the levels x_i and y_i are, in fact, available in an "absolute" sense. By absolute we mean that both calculated and judged levels, in dB, are related to some form of standardized reference pressure. However, although the subjective levels are expressed in terms of the equivalent level of a particular octave band of noise, not all scales are related to any particular definition of perceived level. The loudness and noisiness scales, LL_S, LL_Z, and PNL, are, of course, linked to specific reference sounds, but hazy definitions of these (e.g., the type of random noise, signal level, time histories, listening conditions), together with the practical difficulties of generating those sounds for experimental purposes*, tend to obscure the precise meaning of "absolute" perceived levels. In the case of the weighted sound pressure level scales, no similar subjective definitions exist.

Thus, in order to make the most meaningful comparisons of absolute accuracy, the corrected mean error, Δ, will be utilized. This is similar to $\bar{z}$, except that the subjective level y is expressed in the same units as the calculated level x. Thus, for example, the error Δ for the PNL scale is the mean difference between PNL for the aircraft sounds and PNL for the reference sound. (The actual corrections required for this purpose are given in Table VI.) On the basis of this parameter "accuracy" really pertains to the ability of the scale to rate consistently both aircraft sounds and narrow band noise.

TABLE VI

DIFFERENCE BETWEEN CALCULATED PERCEIVED LEVEL AND OVERALL LEVEL FOR STANDARD REFERENCE SOUND AT L = 100 dB

	PL - L, dB								
	L	L_A	L_B	L_N	L_{NN}	LL_S	LL_Z	PNL	PNL_t
Peak Scale	0	0	0	1.0	0.5	3.0	6.5	5.0	5.0
Effective Scale	-4	-4	-4	-3	-3.5	-1	2.5	1.0	1.0

* Particularly when tape recorders are used. For example, the "octave band" reference sound used in this study had a calculated PNL which was 5 dB greater than its overall level, as analyzed on playback. Maintenance of an adequate signal-to-noise ratio for narrow-band high frequency stimuli is always difficult.

The standard deviation, s, similarly expresses the consistency with which any scale might be expected to rate the relative perceived levels of different aircraft sounds. However, it may not provide a fair test of all scales since it is important to consider the possibility that perceived noisiness "grows" at different rates on the judged and calculated scales, i.e., y is proportional to b.x, where b is a constant other than unity.

4.1.3 Two-variable error analysis. - Although the ideal rating scale should clearly exhibit a one-to-one correspondence with values on the subjective scale, limited departures of the constant of proportionality from a unit value would not detract from the usefulness of any scale which performed consistently since an appropriate slope correction could be applied if necessary.

An alternative to the use of the statistic s is thus to fit the best straight line to the plot of y against x and to measure the dispersion of the data about it. Methods for computing the regression coefficients B_0 and B_1 in the equation

$$y = B_0 + B_1 x \tag{23}$$

are based on minimizing the mean square error of the points $(x_i\ y_i)$ about the line, and can be obtained from any statistics text (e.g., reference 53). It is common practice, for example, for the above equation to represent the regression of y on x in which the error is minimized in the y-direction. Alternatively, a regression of x on y can be performed to determine the coefficients in the equation

$$x = B_0' + B_1' y \tag{24}$$

where the error is minimized in the x-direction. It might be expected that the lines given by equations (23) and (24) are the same so that the slopes B_1 and B_1' are reciprocal. In the event that all the points $(x_i\ y_i)$ lie on a straight line, this is indeed the case; otherwise, as figure 20 (scatter diagrams for particular sets of data) clearly shows, B_1 and $1/B_1'$ differ by an amount which depends on the scatter. The geometric mean of the slopes B_1 and B_1' is called the correlation coefficient R_c where

$$R_c = \sqrt{B_1\ B_1'} \tag{25}$$

This is a useful parameter which describes the correlation between two sets of variables without actually specifying the constant of proportionality. If all the points fall on a line, $R_c = \pm 1$. Otherwise, $|R_c| < 1$ and if the x_i and y_i are completely uncorrelated, $R_c = 0$*. The coefficient can be computed directly from the data by the equation

$$R_c = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\{N \sum x_i^2 - (\sum x_i)^2\} \{N \sum y_i^2 - (\sum y_i)^2\}}} \tag{26}$$

Unfortunately, for the reason that the product $B_1 . B_1'$ is not, in general, equal to unity, the appropriate constant of proportionality linking x and y is not obvious. Reference to figure 20, for example, shows that one slope is high and the other is low relative to the line which would probably be fitted by eye. Although either line is a perfectly valid least squares fit through the data, it seems that one lying somewhere between the two should be used as a general relation relating x and y. Thus, for present purposes, the *geometric mean of B_1 and $1/B_1'$ will be used to indicate the natural slope of the data.* This is defined as $\overline{B}_1$ where

$$\overline{B}_1 = \sqrt{B_1/B_1'} = B_1/R_c \tag{27}$$

These mean lines are also included in figure 20 for comparison. Note that all three lines intersect at the centroid of the data.

Because it is difficult to relate the correlation coefficient to a physical measurement of scatter, the parameter s_{xy}, the standard deviation of the data in the y-direction about the regression of y on x, will also be utilized in subsequent sections. This is sometimes referred to as the "standard error of estimate." Note that it is the scatter about the line with the slope B_1, and not that with the mean slope $\overline{B}_1$. It is given by the equation

$$s_{xy} = \frac{N-1}{N-2} \cdot s_y^2 (1 - R_c^2) \tag{28}$$

where s_y is the total standard deviation of the data in the y-direction.

---

* In practice, a finite value of R_c may be computed, although tests for the significance of its deviation from zero can be applied.

It is, of course, possible that the relationship between x and y is not linear and that errors could be further minimized by fitting higher order curves to the data. However, this possibility has not been investigated in this study, and all analyses have been based upon the assumption of linearity.

4.1.4 Multi-variable analysis. - The possibility that several acoustic and physical variables associated with the aircraft flyover might simultaneously influence perceived level was discussed in section 3.1. This possibility was investigated by a multiple regression analysis. This is the technique by which the joint relationship of a single y-variable upon several x-variables can be established. In its simplest, linear, form, it involves the use of least squares method to fit a multi-dimensional surface of the form

$$y = B_0 + B_1x_1 + B_2x_2 + B_3x_3 + B_4x_4 + \dots \tag{29}$$

to the set of data points. The calculations involved in this procedure are not theoretically complex, but are very lengthy since they involve the solution of as many simultaneous equations as there are x-variables. However, thanks to the electronic computer, the most difficult practical problem is the specification of an appropriate set of x-variables. This is because a successful analysis can only be performed if the x-variables are mutually independent. This is easily explained by a simple example. Suppose we propose signal duration and aircraft distance as two possible x-variables, and it turned out that duration is directly proportional to distance. Then the analysis could have no way of knowing whether or not any apparent dependency upon duration is really induced by the distance variable. In such a situation, a solution becomes indeterminate and the computer analysis fails to complete. Thus, the first step in a multiple regression analysis is to compute the correlation coefficients between each pair of x-variables and to reject all but those which show very small correlations with each other. Unfortunately, for reasons which will be discussed in section 4.2.7, the multiple regression analysis did not reveal any information of value.

4.2 Results

4.2.1 General. - Figure 20 shows all 120 values of judged perceived level (JPL) plotted against the corresponding (a) peak overall levels (L), and (b) the integrated perceived noise levels (EPNL). These objective scales were expected to typify the worst and best scales respectively so that the graphs give some idea of the total range of scatter to be expected. It is certainly most obvious from the figure that the EPNL scale is significantly more consistent than the L scale. It may also be noticed in both plots that one point in each, identified by a different symbol, lies apart from the remainder of the data. These are the two results for sound number 268 identified in Table II as that of a DC-9 on landing approach. A detailed examination of this sound and all analyses

associated with it (see, for example, the data given in Appendix B), reveals no reason why it should appear so different to the remaining data points, particularly those describing similar aircraft and flight conditions. However, because it seems to suffer from a serious, if unexplained error, it has been omitted from the numerical analyses.

The results of the single-and two-variable analyses of the remaining data are presented in Table VII, which lists the various statistics described above for all 18 scales studied. The analysis has been applied to the total set of all 119 sounds, and also to the four subsets of data corresponding to aircraft in the different propulsion system categories. The five parts of Table VII thus correspond to

(a) All 119 sounds (i.e., excluding log 268)

(b) 34 turbojet or turbofan aircraft (jets)

(c) 31 propeller turbine aircraft (turboprops)

(d) 28 piston engined propeller driven aircraft (pistons)

(e) 26 helicopters

An inspection of Table VII reveals that s, the standard deviation of the error $x_i - y_i$ and s_{xy}, the standard error of estimate, which lie in the range 1.8 to 5.3 dB, are generally somewhat larger than the estimated experimental error which lies in the range 1-2 dB. Thus, we can be fairly confident that the differences do, in fact, reflect true differences in the performance of the scales. On the other hand, the total variation of these statistics for any aircraft category perhaps seems rather small, being a mere 65% for the analysis of all sounds for example. These values are, in fact, typical of previous experimental results. However, because in the present study the data sample is very much larger, the confidence with which these differences can be evaluated is very much greater.

A standard method for evaluating the difference between two measurements of data scatter is the F-test. This test utilizes the F-distribution, a mathematically derived function which assigns a probability to the likelihood that the difference in two variances (variance is the square of the standard deviation) occurred by chance. It is based on the assumption that the variances are computed for two samples independently and randomly selected from a normally (Gaussian) distributed population. An inspection of the histograms of the present error distributions gave no reason to suspect that they are not approximately normal. However, it does not seem that the samples can be considered "independent", since each set of objective perceived levels (x-variables) are computed from the same set of 1/3-octave band level arrays. Indeed, the calculated perceived level distributions are highly correlated with each other. Consequently, the validity of the F-test for comparing the variances associated with different scales is somewhat obscure. Nevertheless, there can be no question that the smaller the variance, the more consistent the scale, and there is no reason why the F-test cannot be used as a framework for comparing scales, provided the results are interpreted in a relative sense.

TABLE VII. CORRELATION ANALYSIS.

Category	Statistic*	Perceived Level Scale																	
		L	L_A	L_B	L_N	L_{NN}	LL_S	LL_Z	PNL	PNL_t	EL	EL_A	EL_B	EL_N	EL_{NN}	ELL_S	ELL_Z	EPNL	$EPNL_t$
(a) All 119 Sounds	$\bar{z}$	4.5	-2.1	1.5	1.8	3.5	7.5	9.5	9.8	13.0	0	-6.7	-3.2	-2.3	-1.9	3.0	5.2	4.9	7.9
	Δ	4.5	-2.1	1.5	0.8	3.0	4.5	3.0	3.8	7.0	4.0	-2.7	0.8	0.7	1.6	4.0	2.7	3.9	6.9
	s	4.4	3.9	4.2	4.1	4.9	3.7	3.0	3.9	4.5	4.0	3.6	3.5	2.8	3.3	2.8	2.7	2.8	3.2
	R_c	.822	.881	.855	.881	.898	.907	.912	.905	.876	.821	.871	.869	.929	.936	.917	.913	.930	.910
	B_1	.695	.707	.692	.684	.597	.704	.817	.693	.644	.797	.787	.809	.817	.728	.853	.939	.813	.789
	s_{xy}	3.7	3.1	3.4	3.1	2.9	2.8	2.7	2.8	3.2	3.8	3.2	3.3	2.4	2.3	2.6	2.7	2.4	2.7
	$\bar{B}_1$	.845	.802	.809	.776	.665	.775	.885	.766	.735	.972	.904	.931	.879	.778	.930	1.027	.875	.868
(b) 34 Jet Aircraft Sounds	$\bar{z}$	2.5	-2.4	0.3	2.3	4.7	6.8	9.5	9.4	12.1	-0.6	-6.3	-3.2	-1.7	-1.0	3.1	5.6	5.3	7.6
	Δ	2.5	-2.4	0.3	1.3	4.2	3.8	3.5	4.4	7.1	3.4	-2.3	0.8	1.3	2.5	4.1	3.1	4.3	6.6
	s	3.5	2.7	3.5	2.7	4.3	2.9	2.2	2.8	3.3	4.1	2.9	3.8	2.4	2.5	2.5	2.8	2.4	2.3
	R_c	.891	.935	.895	.952	.927	.947	.951	.954	.940	.820	.912	.849	.941	.961	.938	.921	.943	.947
	B_1	.804	.851	.787	.798	.654	.785	.960	.784	.742	.871	.987	.898	.931	.811	.942	1.112	.913	.915
	s_{xy}	3.2	2.5	3.2	2.2	2.7	2.3	2.2	2.1	2.4	4.1	2.9	3.7	2.4	2.0	2.5	2.8	2.4	2.3
	$\bar{B}_1$	.902	.910	.880	.838	.705	.829	1.009	.822	.789	1.062	1.082	1.059	.989	.844	1.004	1.218	.968	.966
(c) 31 Turbo-prop Aircraft Sounds	$\bar{z}$	6.4	-0.5	3.4	4.5	4.8	9.2	10.0	11.5	15.0	-0.1	-6.9	-3.2	-2.0	-1.3	3.2	4.9	5.2	8.4
	Δ	6.4	-0.5	3.4	3.5	4.3	6.2	3.5	6.5	10.0	3.9	-2.9	0.8	1.0	2.2	4.2	2.4	4.2	7.4
	s	4.5	4.3	4.7	4.0	5.1	3.7	3.4	4.0	5.1	4.3	3.6	3.9	2.8	3.3	2.6	2.7	2.7	3.5
	R_c	.810	.839	.827	.890	.872	.910	.875	.897	.864	.784	.849	.838	.921	.915	.925	.901	.925	.904
	B_1	.670	.674	.629	.666	.577	.686	.776	.662	.578	.724	.814	.738	.800	.729	.850	.983	.817	.725
	s_{xy}	3.8	3.5	3.6	2.9	3.1	2.7	3.1	2.8	3.2	4.0	3.4	3.5	2.5	2.6	2.4	2.8	2.4	2.7
	$\bar{B}_1$	.827	.803	.761	.748	.661	.754	.887	.738	.669	.923	.959	.881	.869	.796	.919	1.091	.883	.802
(d) (28) Piston Aircraft Sounds	$\bar{z}$	6.0	-1.8	2.7	-0.1	3.9	8.5	10.1	10.7	14.3	0.4	-7.7	-3.0	-2.6	-2.4	2.7	4.6	4.9	7.4
	Δ	6.0	-1.8	2.7	-1.1	3.4	5.5	3.6	5.7	9.3	4.4	-3.7	1.0	0.4	1.1	3.7	2.1	3.9	6.4
	s	3.0	3.1	2.6	3.4	4.4	2.9	2.2	2.9	3.3	2.9	3.5	2.5	2.4	3.5	2.4	2.1	2.5	2.8
	R_c	.927	.948	.955	.911	.948	.965	.967	.964	.935	.922	.930	.953	.969	.969	.967	.966	.972	.959
	B_1	.855	.768	.835	.804	.646	.764	.855	.762	.761	.907	.746	.864	.815	.695	.814	.873	.794	.783
	s_{xy}	2.8	2.3	2.2	3.0	2.3	1.9	2.9	2.0	2.6	2.9	2.7	2.2	1.8	1.8	1.9	1.9	1.8	2.1
	$\bar{B}_1$	.922	.810	.874	.883	.682	.792	.884	.790	.814	.984	.802	.907	.841	.718	.842	.904	.817	.816
(e) (26) Heli-copter Sounds	$\bar{z}$	3.1	-3.8	-0.3	0	0	5.2	8.2	7.5	10.3	0.5	-6.5	-3.3	-2.9	-3.1	2.6	5.5	4.3	7.0
	Δ	3.1	-3.8	-0.3	-1	-0.5	2.2	1.7	2.5	5.3	4.5	-2.5	0.7	0.1	0.4	3.6	3.0	3.3	6.0
	s	5.3	4.9	4.8	4.6	4.6	4.2	3.8	4.6	4.6	4.5	4.2	3.8	3.4	3.5	3.3	3.1	3.5	3.6
	R_c	.653	.739	.716	.743	.724	.731	.749	.736	.735	.649	.726	.724	.777	.780	.745	.757	.766	.755
	B_1	.410	.449	.458	.481	.478	.514	.566	.473	.475	.482	.521	.571	.620	.614	.681	.743	.605	.593
	s_{xy}	3.4	3.0	3.1	3.0	3.1	3.0	2.9	3.0	3.0	3.4	3.1	3.1	2.8	2.8	3.0	2.9	2.9	2.9
	$\bar{B}_1$	.628	.608	.625	.647	.661	.703	.756	.643	.646	.743	.718	.789	.798	.787	.914	.982	.790	.785

*x_i = calculated perceived level of i-th sound

y_i = sound pressure level of reference sound judged equally noisy

$\bar{z}$ = mean value of error $x_i - y_i$

s = standard deviation of $x_i - y_i$

R_c = correlation coefficient (between x_i and y_i)

B_1 = slope of line with minimum error in y-direction

s_{xy} = standard deviation of points about above line

$\bar{B}_1$ = best mean slope = B_1/R_c

y_i' = calculated perceived level of reference sound judged equally noisy

Δ = mean value of $x_i - y_i'$

A description of the F-test may be found in any statistics text, (e.g., reference 59) which will also tabulate critical values of the F-distribution. Provided the normality assumption holds, these tables may be consulted to determine, with some specified degree of confidence, whether or not we can reject the probability that the differences between two variances occurred by chance. Very simply, the hypothesis is tested that the F-ratio, formed by dividing the larger by the smaller of the two variances, is significantly different to 1.0. For equal samples, the larger the number of data points the smaller the critical F-ratio. For example, if the number of data points in each sample is 10, the ratio of their two variances would have to exceed 3.18 for us to be 95% confident that the two samples were not drawn from the same sample. If the samples contain 119 measurements, then the critical ratio is reduced to 1.35.

Although, for the reasons outlined above, we cannot be sure that these specific numerical values are valid for our purposes, the fact that the differences between variances should only be assessed with regard to sample size is a very important one. Accordingly, figure 22 has been prepared to summarize the standard deviation data presented in Table VII. The logarithmic vertical scale in this figure represents the standard deviations s or s_{xy} in dB. Constant linear increments on this scale separate standard deviations of a constant F-ratio. Consequently, these increments effectively represent bands within any of which the differences have equal probability of occurring by chance.

For each of the data sets, the objective noise rating scales have been ranked in ascending order of both s and s_{xy} by plotting their positions on this logarithmic scale. Each column has also been divided up into segments corresponding to the critical F-ratios for the 5% level of significance (95% confidence). Thus, the rating scales appearing within any particular segment are statistically equivalent within some confidence limits. Note also that the segments have been arbitrarily positioned, vertically. The zones are strictly free to slide up and down, and they have been denoted merely to show their variable height from set to set.

Confining attention initially to the results in Table VII(a) for the complete 119 aircraft set two facts are immediately apparent. The first is that the errors s_{xy} about the regression lines are substantially less than the standard deviations s, leading to a significant improvement in consistency when a "floating slope" is allowed. Secondly, the integrated duration corrected scales, the "effective" perceived levels, are significantly more consistent than the peak levels.

With regard to the first of these, it seems that the slope of the line is of first order importance to the entire problem of aircraft noise rating. For the complete data set, mean slopes ($\bar{B}_1$) between 0.665 (for L_{NN}) and 1.027 (for ELL_Z) have been computed, and it is necessary to examine possible explanations for this in some detail. The slope $\bar{B}_1$ compares the average rate of growth of judged level to that of calculated perceived level over the experimental sound pressure level range (from 84 to 115 dB, peak). Here, the

judged perceived level is the sound pressure level of the octave band standard reference, which is judged equal in perceived magnitude. Since for the L scale, $\bar{B}_1 = 0.845$, we see that the actual perceived level is proportional to 0.845 x the peak intensity of the aircraft sounds. In other words, the perceived level of aircraft noise does not grow with overall sound pressure level at the same rate as that of the octave band of noise at 1000 Hz. We saw in section 2.1 (figure 11) that for wideband spectra, this effect is, in fact, predicted by the loudness scale LL_Z for all levels above approximately 70 phons and below 100 phons by LL_S. However, PNL and the sound pressure level scales predict equal growth (for relatively non-changing spectra) at all moderate and high levels. It is not surprising, therefore, that LL_Z yields the highest slope ($\bar{B}_1 = 0.885$), indicating that this scale, for aircraft noise spectra as well as the idealized spectrum used for demonstration purposes in section 2.1, predicts a lower growth of perceived level with intensity than do the other methods.

A major influence upon the results of this study may have been the decision to play sounds to the subjects at realistic levels, i.e., as close as possible to the original levels at the time of recording. This maintained realistic relationships between overall signal intensity, spectral content and sound duration, which appear to have an important bearing upon the practical applicability of the noise rating methods.

In the first place, ignoring variations in aircraft size and power, an increase in level implies a reduction in distance between the source and the observer. This, in turn, is accompanied by a reduction in signal duration, and because of reductions in atmospheric sound absorption at higher frequencies, an increase in the high frequency content. These trends may be observed by scanning through the figures in Appendix B. The level-duration relationship was, in fact, investigated by computing the correlation between PNL (x-variable) and its "duration correction," which is equal to EPNL-PNL (y-variable). The correlation coefficient R_c and slope $\bar{B}_1$, were, respectively, -0.474 and -.133, showing the expected reduction in duration with increase of signal level and a significant correlation. The frequency effect can be inferred by comparing results for the scales L and L_{NN}. The latter gives substantially more weight to high frequencies and less to the low frequencies than a linear weighting function, so that the difference between the two levels gives a good indication of the distribution of energy between low and high frequencies. Since the slope $\bar{B}_1$ is substantially less for L_{NN} (0.665) than for L (0.845), it is clear that the difference between L_{NN} and L increases fairly rapidly with level, indicating a shift of emphasis to higher frequencies as intensity increases. It is interesting that scales which give increasing emphasis to high frequencies ($L \rightarrow L_A \rightarrow L_N \rightarrow L_{NN}$) progress to ever decreasing slopes ($0.845 \rightarrow 0.802 \rightarrow 0.776 \rightarrow 0.665$).

Because of the level-duration relationship, the application of a duration correction causes an increase of slope, which may be seen for all the effective scales listed in Table VII(a). In particular, ELL_Z exhibits a constant of proportionality which is very close to unity. Also, the integration has caused an improvement of consistency in many

of the scales. Comparing s_{xy} values, this is particularly noticeable in EL_{NN}, EL_N, EPNL and $EPNL_t$. These procedures are the ones which are most sensitive to high frequencies and this observation thus suggests that the duration correction (negative) is tending to compensate for what is perhaps an excessive emphasis upon high frequency content (positive).

Note that just as s and s_{xy} tend to equalize as $\overline{B}_1$ approaches unity (not exactly true because of the different denominators in equation (22) and (28)), s shows a marked improvement from the "peak" to the "effective" scales.

Comparing the various procedures in figure 22, it appears that if importance is attached to unit slope, the three effective perceived level scales ELL_S, ELL_Z and EPNL, together with EL_N and LL_Z, are significantly better than the remainder. Reasons for this can be traced to either high slope (LL_Z, ELL_Z), low scatter (EPNL, EL_N) or both (ELL_S). If, on the other hand, slope is ignored, we see by comparing the standard errors of estimate s_{xy} that the same procedures rank highly, but that all have been overtaken by EL_{NN} with a very low error of 2.3 dB. It seems that emphasis upon high frequencies markedly improves consistency, but that because of the particular relationship between intensity and frequency distribution, this step has caused an excessive increase in the calculated perceived level growth rate ($\overline{B}_1 = 0.778$).

Moving down the rank list (under s) we see that the next group of scales includes the remainder of the duration corrected versions (except EL). Significantly lower again are the peak level scales LL_S, PNL, L_N, L_A and finally, at the bottom of the list, L, L_B, L_{NN} and PNL_t. For exactly opposite reasons for which the best scales are superior, the poorest ones have low slopes, high scatter, or a combination of both. It will be noticed, for example, that in the s_{xy} column, the peak scales have exchanged places with the effective scales, so that the uncorrected versions of the more elaborate scales remain superior to the corrected versions of the poor sound pressure level scales.

Turning now to the question of accuracy, which has been related to the ability of the rating procedures to accurately scale both narrow and broadband noise, the mean errors $\overline{z}$ and Δ for the 119 sound set are compared graphically in figure 23. It will be remembered that $\overline{z}$ is simply the mean difference between the calculated levels and the average sound pressure levels of the equivalent standard reference sounds, whereas the increment Δ is based upon the calculated perceived level of the reference. The increment $\overline{z}$ thus gives the direct differences between mean levels calculated on the different scales. We see, for example, a well known increment of 8 dB between L_N and PNL. Also note that LL_Z generates levels which are 2 phons higher than LL_S, but that LL_Z and PNL are very close.

These differences reveal the different ways in which the scales account for the increase in perceived level with bandwidth, remembering that for a narrow band signal centered at 1000 Hz, the levels would all agree to within 1 dB. Confining attention to the more meaningful increment Δ, it seems that for the average aircraft flyover sound, the complex loudness/noisiness procedures, with or without duration allowances, overestimate perceived level (with respect to that of a 1000 Hz reference) by around 4 dB. The tone correction increases this discrepancy by a further 3 dB. The sound pressure level errors, on the other hand, range between +4.5 dB and -2.7 dB, in each case reflecting the net attenuation introduced by the weighting function. Thus, the linear scales L and EL overestimate by 4.5 dB and 4.0 respectively, whereas L_A and EL_A underestimate by 2.1 and 2.7 dB.

4.2.2 Differences between aircraft categories. - Apparently confirming the significance of subjective differences in the acoustic characteristics of aircraft with different propulsion systems, figures 21 and 22 reveal that clear differences do indeed exist between the consistency of the scales as applied to the different data sets. Unfortunately, because of the smaller samples, distinctions between scales are less clear but it is obvious that, on an average, the scales are most consistent for the piston sounds followed by the jets, the turboprops and the helicopters, in that order.

The results for the helicopters are remarkable in that (a) all scales are poor, and (b) in terms of standard error of estimate s_{xy}, there is practically no difference between any of the scales. Although L and EL appear inferior, the differences are not significant. However, reference to the standard deviation s does help to discriminate between the scales because there is a large variation of $\overline{B}_1$ (Table VII(e)). On the scale of s, the methods can be divided into two basic categories, moderate and poor, with the effective perceived level scales being superior to a group containing all the peak scales plus EL, EL_A, and EL_B. The duration correction is particularly beneficial, probably because of the long durations associated with some of the very low speed flyovers. The reason for the consistently poor performance of the scales is probably related to the domination of the helicopter sounds by low frequency energy of a pulsatile nature. We have seen that attention to high frequencies is one of the major factors which, in general, discriminates between the better and poorer scales. The fact that the helicopter sounds contain little high frequency energy, therefore, serves to explain the tight concentration of the scales in the s_{xy} column. The fact that all scales are poor suggests that the subjective effects of low frequency pulsatile sounds require further investigation.

Proceeding to the other end of the range, most of the better scales perform remarkably consistently for the sounds of piston engined aircraft, with s_{xy} errors of only 1.8 dB which are probably as low as possible. However, the slopes $\overline{B}_1$ are consistently low, and further, they are not significantly increased by the application of a duration correction.

This is possibly because sound pressure level is more strongly related to aircraft size than to distance for the piston group. Because the slopes are low, the deviations s are substantially bigger than the deviations about regression. Of particular interest is that although the duration correction has a very marked effect upon L_N and L_{NN}, it does not improve LL_S or LL_Z in terms of s_{xy}. Again, the only explanation which can be offered is that the (negative) duration correction is counteracting some harmful effect of emphasizing the high frequencies (which LL_S and LL_Z do to a much lesser extent than either L_N, L_{NN} or, for that matter, PNL).

Figure 24 attempts to illustrate typical spectra for the different aircraft categories. Reference to Appendix B will show that these are at best very liberal generalizations since variations about them are very large. However, it is believed that they do portray the significant differences between the categories. It should be pointed out that it has been found impossible to detect any such categorical differences between the spectra of the turboprops and the pistons, with the possible exception of occasional high frequency spikes in the former, which are absent in the latter. For this reason and because the scales are considerably less consistent for the turboprops than for the pistons, as figure 22 clearly shows, it can only be concluded that the rating procedures are failing to account for some key feature of the turboprop sounds. This will be discussed further in paragraph 4.2.6.

The most notable feature of the results for the 34 jet sounds, Table VII(b), is that the average slope of the regression lines is rather greater than it is for the other data sets. In fact, for the effective scales, the average slope is very near to unity and LL_Z here results in a $\overline{B}_1$ of 1.218. This says that perceived level grows more rapidly with intensity for jets than it does for other aircraft. Like the sounds of other aircraft, and perhaps moreso since the total power range is somewhat smaller, signal level is closely linked with aircraft proximity, and therefore, with signal duration and frequency distribution. Also, the higher frequency energy appears in the form of compressor, fan or turbine tones. Thus, it must be conjectured that the presence of these components at the higher sound pressure levels is responsible for the relatively high growth of judged level. It is certainly not without significance that $EPNL_t$, with the tone corrections, appears high in the s column for jets. In general, the scales PNL and L_N fare particularly well, scales which were previously noted to require a duration correction to compensate for possible overemphasis upon high frequencies.

In terms of absolute accuracy, an inspection of Table VII shows that the mean error Δ for any scale does not vary significantly between aircraft categories.

Based on the observations discussed in this section, and upon others to follow, the major factors in perceived level estimation, frequency weighting, band summation procedures, duration corrections, tone corrections, together with other factors of possible importance, will now be discussed separately in turn.

4.2.3 Band level summation procedures. - Because both the band level summation technique and the frequency weighting function contribute to the performance of any particular scale it is difficult to isolate the effects of either one. The fact that the five sound pressure level scales differ only in the form of their frequency weighting functions allows some conclusions to be drawn regarding the independent effects of the latter and these are discussed in the next section. A significant result regarding band summation may be found in Table VII(a) which shows that the two scales EL_N and EPNL are practically identical in every respect except in the mean errors $\bar{z}$ and Δ. Both scales overestimate the judged level of the aircraft sounds but the error for the EL_N (Δ= 0.7 dB) is rather less than that for EPNL (Δ= 3.9 dB). Since these scales utilize practically identical frequency weighting functions, this finding provides an important comparison between the different band level summation procedures incorporated in the two methods.

Reference to figure 9 shows that EPNL may be expected to exceed EL_N when the effective number of 1/3-octave bands in the signal exceeds 6 or so. To shed some further light on this, a further analysis was made of the Perceived Noise Level computations which involve the Stevens summation rule

$$N_t = N_{max} + F(\sum N_k - N_{max}) \tag{30}$$

where N_k is the noisiness of the k-th 1/3-octave band in noys, N_{max} is the noy value for the noisiest band, F is a constant (0.15) and N_t is the effective noisiness of of the total complex signal. The summation $\sum$ is performed over all bands. The analysis showed that, on average, for all sounds,

$$\begin{aligned} N_{max} &= 0.4\, N_t^{*} \quad \text{(standard deviation - } N_t) \\ \text{whence} \qquad \sum N &= 4.5\, N_t \\ &= 11\, N_{max} \end{aligned} \tag{31}$$

---

* The corresponding fractions for the jet, turboprop, pistons and helicopter subdivisions were, respectively, 0.38, 0.46, 0.39 and 0.37.

Thus if we revert to our concept of "uniformly distributed noise" in which all the N_k are equal (and equal to N_{max}), such a signal would contain eleven effective 1/3-octave bands of noise. Note that figure 9 shows a difference of 3 dB between the PNL and sound pressure level curves for $K = 11$.

Both the 4 dB discrepancy by which EPNL overestimates the average perceived level of aircraft noise and the sub-unity slope of 0.875 can be corrected by introducing a variable factor F' into equation (30), as recently proposed by Stevens (ref. 2), such that

$$N'_t = N_{max} + F' \left(\sum N_k - N_{max} \right) \tag{32}$$

Since this formula must reduce the perceived level by 4 PNdB the conversion formula (20) gives the result

$$33.3 \log_{10} (N_t/N'_t) = 4$$

or (33)

$$(N_t/N'_t) = 1.32$$

i.e.

$$\frac{N_{max} + 0.15 \left(\sum N - N_m \right)}{N_{max} + F' \left(\sum N - N_m \right)} = 1.32 \tag{34}$$

For a summation over 11 effective bands this yields

$$F' = 0.09$$

This value applies at the average level (which occurs at JPL = 96 dB) where PNL ~ 100 PNdB. At this level the effective perceived noisiness N_t is 64 noys so that, from equation (31), $\sum N \sim 290$ noys for the average aircraft sound. Accordingly, we

may write the relationship for a unit $\bar{B}_1$ slope:

$$\log_{10} N \left\{1 + F'(K - 1)\right\} = 0.875 \log_{10} N \left\{1 + 0.15(K - 1)\right\} \qquad (35)$$

where N is the effective (uniform) band noisiness and $K = 11$. Whence we obtain

$$F' = 0.4(\sum N)^{-1/8} - 0.1 \qquad (36)$$

This function, which decreases with signal intensity, is plotted in figure 25 and is compared with Stevens' MK VII recommendation and the function derived from Zwicker's charts in figure 10. It is seen to be significantly smaller than both. This rather low value for F', which only reaches the value 0.15 at approximately 70 PNdB, must be assumed to be characteristic of aircraft noise spectra and of course, is only known to be applicable for Effective Perceived Noise Levels in excess of 85 EPNdB. The growth of computed perceived level with bandwidth for $F = 0.09$ has been included for comparison with larger values in figure 9. As expected, this curve is 1 dB lower than the sound pressure level curve at $K = 11$ and is 6 dB lower than the result for $F = 0.15$ for high K. For practical values of K, however, it is unlikely that the $F = 0.09$ curve will differ from the sound pressure level curve by more than one or two decibels. This result would seem to have considerable practical significance since it strongly suggests that, at least for aircraft noise in the mid-level range, the complex noisiness calculation procedure, even in its modified form, may be accurately approximated by the N-weighted sound pressure level scale.

It may be noted that although figure 9 suggests that Zwicker's Loudness Scale, LL_Z, will estimate levels around 6 dB greater than LL_S or PNL, this increment is not evidenced in Table VII. Indeed, LL_Z is lower by approximately 1 dB. This may be attributed to two factors. The first is that the curve in figure 9 corresponds to a signal level of approximately 80 phons where the masking effect is reduced so that individual band contributions are greater. (At higher levels the bandwidth effect is less marked). The second is that the neglect of the masking profiles in the computations may cause larger errors than were originally anticipated.

4.2.4 Frequency weighting functions. - The perceived level calculation procedures evaluated in this study utilize a very wide variety of actual or effective frequency weighting functions, which are illustrated in figure 2. These are based upon different

experimental measurements of equal perceived magnitude functions and the variety of scales reflects the use of different experimental conditions, methods and environments. The curve derived from Zwicker's data for the LL_Z procedure was not measured directly; it is based on measurements of critical bandwidths, masking, and head diffraction, and was obtained by computation from Zwicker's loudness evaluation charts (ref. 18). It has been normalized to zero attenuation at 1000 Hz for direct comparison with the other curves. The NN-curve, although obviously indicating much greater sensitivity at higher frequencies than do the other curves, has been included because (a) it was measured in the same test chamber in which the present experiments were performed, and (b) it does provide an extreme case for study.

Because of different band summation procedures utilized, it is difficult to compare the sound pressure level weightings with the functions included in the perceived level procedures PNL, LL_S, and LL_Z. Insofar as these complex scales are concerned, assuming that the main summation differences lie in the growth functions discussed in the previous sections, their different frequency functions can be compared through the statistic s_{xy}. Based on overall performance as indicated by Table VII(a), EPNL appears a little more consistent than ELL_S, which in turn is a little better than ELL_Z, although none of the differences appear significant. Reference to figure 2 shows that this reflects the order of decreasing emphasis upon high frequencies.

In the case of the five sound pressure level scales, however, the relative merits of the weighting functions (Linear, A, B, N, and NN) can be compared directly, since the same band summation method is common to each scale. This comparison, however, must take into account the three related effects upon the constant of proportionality $\overline{B}_1$, the mean error Δ , and the scatter, as reflected by s_{xy}. Trends in all three quantities may be associated with the degree of emphasis upon high frequencies. Specifically, as this emphasis is increased (from EL to EL_B to EL_A to EL_N to EL_{NN}) the slope $\overline{B}_1$ decreases from .972 to .778, the deviation s_{xy} decreases from 3.8 to 2.3 and the mean error Δ tends to reduce although this is primarily related to the net attenuation by the weighting networks.

The high sensitivity of the slopes $\overline{B}_1$ to the frequency parameter is a function of the fact that for the particular data studied, level and frequency content are correlated via the distance variable. Thus, high frequency emphasis causes greater calculated level increases at higher signal intensities than it does at lower levels, effectively introducing a change of slope. Although this change leads to higher correlation, this benefit is offset by the reduced slope. An optimum weighting function can only be defined therefore in terms of a tradeoff between scatter, slope and mean error. Certainly, of those studied, the N-network appears to most closely approach the ideal. When used with a duration correction, EL_N yields a mean error $\Delta = 0.7$ dB, a standard error of estimated $s_{xy} = 2.4$ dB and a mean slope $\overline{B}_1 = 0.879$ based on the complete set of data. Although improvements are undoubtedly possible by careful attention to detailed network design, this performance is certainly good and in fact slightly better than that of EPNL.

4.2.5 Duration effects. - The results clearly show that the integrated duration correction has a beneficial effect upon the performance of the scales, both in terms of consistency and slope. An approximate correction, based on the actual time between the 10 dB-down points, was included in the STOL noise study (ref. 10), but proved significantly inferior to the integrated version. It was, therefore, omitted from the present analysis.

In order to examine the duration effect a little more closely, the correlation was computed between the subjective levels and levels calculated according to the relationship

$$EPNL' = PNL + K(EPNL - PNL) \tag{37}$$

where the constant K was varied over an appropriate range. This is equivalent to the equation

$$EPNL' = PNL + 10K \log_{10} T_e \tag{38}$$

where T_e is the effective duration computed by integration. The analysis was applied to the five different sets of data and the results are shown in figure 26. This illustrates, for each group, the variation of s and s_{xy} with K.

In all cases, the minimum in the s-curve occur near K = 1, suggesting that the constant of 10 is indeed optimum in all cases. However, it should be remembered that because of the low $\bar{B}_1$ slopes encountered, the scatter s is reduced by the increase of slope which the duration correction brings about. The curves would be somewhat different if the growth problem were remedied as discussed in section 4.2.2 and would, in fact, bear more resemblance to the s_{xy} curves. It may be seen that the troughs in these curves, in addition to being shallower, tend to occur at fractional values of K. This is particularly noticeable in the case of the jet data. Nevertheless, for the complete set of 119 sounds the minimum occurs at K = 0.9 and because s_{xy} is not particularly sensitive to K anyway, the presently used 3 dB per time doubling seems to be a good choice, at least for the PNL scale.

4.2.6 Tone correction. - The same cannot be said of the tone correction which has not shown itself to be a particularly beneficial measure since, in general, its application has caused both PNL and EPNL to become less consistent evaluators of perceived level. The probable reasons for this, however, may be identified by inspection of the results for the different aircraft categories.

It does seem significant, for example, that the one case in which the tone correction proves advantageous is the application of $EPNL_t$ to the jet sounds. High fre-

quency tones could be observed in 26 of the 34 sounds used, and were "strong" in about 10 of these cases. However, tone corrections of between 1 and 5 dB were applied to all sounds, without exception, and the average increment was 2.3 dB. In order to determine whether or not these corrections could be related to the subjective intensity of the tone, an attempt was made by the author to assign a numerical value between 0 and 4 to the judged "strength" for each sound. The correlation coefficient between the tone correction increment and this parameter was computed to be 0.30, a fairly low, but nevertheless significant correlation for this very crude experiment.

It was initially somewhat surprising to find that even larger corrections with mean values of 3.2 and 2.5 dB were applied to the turboprop and the piston engine data respectively, although an inspection of the spectra in figure 30 quickly reveals why. Very large spikes occur at the fundamental propeller frequency, normally in the region of 100 Hz, and sometimes at its higher harmonics. Thus, even though no high frequency tone was present in the case of the piston sounds, corrections as high as 5 dB were automatically applied by the $EPNL_t$ procedure. These spikes do, of course, correspond to "tones" in the spectrum. However, the quality of propeller sounds is controlled more by the higher harmonics than the fundamental. In fact, it is a well known fact that the "impulsiveness" of propeller noise, and harmonic sound in general, increases as the spectrum becomes more flat. It is thus conceivable that the tone correction, as presently constituted, works "in reverse" for the sounds of propeller aircraft, adding larger tone corrections as the harshness of the sound decreases.

However, obvious differences between the piston and turboprop groups do exist, as discussed in section 4.2.2, and, for the present, these can only be attributed to the presence of compressor and fan components in the case of the turboprops and perhaps exhaust components for the pistons. For some reason all scales, with and without tone corrections, are less consistent for the turboprops and it is possible that the tone correction might have proved advantageous for this group, as in the case of the jets, if it could have operated upon the high frequency tones only.

In view of the uncertainties regarding the subjective aspects of propeller and rotor noise, it would seem adviseable for the moment to restrict the tone correction to higher frequencies, say above 500 Hz, and to ignore "tones" identified at lower frequencies.

4.2.7 Other non-auditory factors. - In an attempt to isolate possible influences of "non-auditory" factors on subjective response, a multiple regression analysis was performed to establish the dependence of judged level on various combinations of variables selected from the following list:

Calculated perceived level*
10-dB down duration*
Aircraft Power (or thrust)
Aircraft gross weight
Number of engines
Minimum distance between aircraft and microphone
Estimated aircraft velocity
Aircraft category

No significant dependencies could be found, except upon the first two variables and those which could be explained through the high correlation between certain variables from this list and the acoustic variables accounted for by the various noise rating procedures. It is concluded that because of high inter-variable correlations, it is not possible to derive useful results by this technique with the available sample sizes.

* Several of the noise rating procedures were used as the basis for the first two variables in this list.

5.0 CONCLUSIONS

A large scale experiment has been performed to determine the practical value of existing noise rating procedures for application to aircraft flyover noise. Up to 32 subjects took part in laboratory paired comparison tests to judge the perceived levels of 120 different aircraft flyover recordings. The results were analyzed to test the performances of eighteen noise rating methods which were selected to represent a wide range of possible alternatives. The objectives of the study were to determine (a) whether the reputed "better" methods, which have been developed and tested in a series of rather limited experiments, live up to their reputations in the widest possible range of practical applications, or whether as has often been suggested, there is little practical difference between a multitude of alternative choices; (b) whether non-auditory information conveyed by the sound, such as aircraft type, distance and speed, influence subjective response; and (c) to recommend possible refinements to the scales, if any, which might improve their performance.

The 120 sounds represent an unusually large sample for tests of this nature and were selected to cover the widest possible range of aircraft types, operating conditions, dynamic range and signal duration. Also, the sounds were divisible into very roughly equal samples in each of the four categories: turbojet (or fan) powered aircraft, propeller-turbine aircraft, piston engined aircraft, and helicopters. This allowed meaningful analyses to be applied to each. The various noise rating procedures were evaluated in terms of their ability to accurately and consistently predict the perceived levels of the sounds as judged by comparison with a standard reference of an octave band of noise centered at 1000 Hz. The experimental error was estimated to be sufficiently low (1-2 dB), and the sample size sufficiently large, that the differences between scales could be distinguished with an unprecedented degree of confidence. The conclusions derived from the investigation are as follows:

1. It was found that significant differences do exist between scales and that they can be ranked into several strata. In terms of consistency, the better methods are essentially indistinguishable and include the three "complex" perceived level procedures ELL_S, ELL_Z and EPNL due to Stevens, Zwicker, and Kryter, and where the prefix E denotes the application of an integrated signal duration allowance. Also statistically indistinguishable from these for the aircraft sounds, were integrated or "effective" sound pressure level measured on the N-scale, EL_N, and peak loudness level, LL_Z, using Zwicker's method.

   However, with the exception of Zwicker's procedure, all methods tend to overestimate the growth of perceived level with intensity over the range of sound pressure levels investigated (84-115 dB overall).

2. Distinct differences were observed between the applicability of the scales to sounds in the four different aircraft categories. On average, the scales were extremely consistent for the piston engined aircraft sounds but increasingly less so for the jets, the turboprops and the helicopters, in that order. The deficiencies in the latter groups are attributed to improper account of pure tones in the turboprop spectra and low frequency harmonic sound in the case of the helicopters.

In an attempt to identify precise reasons for the variations in performance, a number of basic factors were investigated. The conclusions related to each are described below:

3. Band level summation procedures. - Of the three alternative approaches, Zwicker's is the most soundly based upon auditory theory, best explains the experimental observations relating to the growth of perceived levels, and possibly takes automatic account of spectral spikes. Stevens' technique, which is common to both LL_S and PNL, is based upon a more empirical model but for practical, wideband sounds it turns out to be remarkably similar to Zwicker's calculation at lower sound pressure levels. However, both techniques overestimate the perceived level of aircraft noise with respect to that of the 1000 Hz reference by an average of 4 dB. At higher levels the Zwicker and Stevens procedures differ in that the Stevens' approach overestimates the growth of perceived level of wideband sounds with respect to that of narrowband sound. Based upon the experimental findings and an investigation of the relationships between the three basic summation procedures, a simple remedy for both problems has been defined for use in the EPNL calculation. This involves a smaller, but variable F-factor in the loudness/noisiness summation formula.

The simple energy summation process performed by the weighted sound pressure level circuits is rather sensitive to the particular choice of weighting network and, depending on this selection, can over- or underestimate the perceived level of wideband noise relative to that of narrowband noise. Thus, a linear (flat) function overestimates, the A-weighting underestimates, whereas the N-weighting, based on the inverse of the 40 noy contour, shows a very small mean error. Otherwise the energy summation rule gives a very good approximation to the revised noisiness summation rule over a practical bandwidth range.

4. Frequency weighting. - The procedures LL_S, LL and PNL directly or indirectly utilize similar frequency weighting functions and largely for this reason tend to be equally consistent. An investigation of a set of widely differing sound pressure level weighting functions revealed an improvement in performance as emphasis shifted from low frequencies to high. However, on the basis of consistency, perceived level growth and accuracy, the N-weighting is the best of those studied and is probably close to optimum.

5. Duration correction. - Based on the assumption of a uniform duration/perceived level tradeoff allowance, the presently used correction of 3 dB per duration doubling is close to optimum for aircraft sounds in all categories. The application of this duration allowance improves the performance of the scales.

6. Tone corrections. - The FAA-Little tone correction to the PNL procedure was tested for each aircraft category. Only in the case of the jet sounds did the correction appear to perform as intended, and then the improvement was marginal, a slight improvement in consistency being offset by a further increase in the mean error. It was concluded that in the case of the piston sounds the correction was not required and that for the turboprops the need for a correction possibly exists, but that this need was not fulfilled by the selected procedure. The problem appears to lie not entirely with the form or magnitude of the correction, but in the manner by which "tones" in the spectra are detected by the computer model.

7. Other non-auditory factors. - No correlation could be found between judged perceived level of the sounds and the non-auditory variables defining size, power, distance or velocity. It is concluded that whether or not such influences exist, they are practically impossible to detect because of their high correlation with auditory variables.

In summary, the search for improved methods for rating aircraft flyover noise does appear worthwhile and indeed, it is possible that recent advances have already improved upon the methods investigated herein. It is further considered that a sufficiently large body of data now exists to enable significant advances to be made without recourse to further experimentation. It is therefore recommended that effort be concentrated upon using this data to:

1. Establish an optimum weighting function for use in a standardized sound pressure level scale. It is probable that such a scale, used with a time-integration, could prove equal to any possible alternative for practical evaluation of aircraft noise, particularly for noise monitoring.

2. Further refine the perceived level procedures due to Stevens (including PNL) and Zwicker. Specifically, it is likely that Zwicker's method could be improved by revision of the audibility threshold and head diffraction functions, and that the LL_S/PNL methods could benefit from both the research performed under (1) and further attention to the loudness growth function.

3. Develop an improved tone identification mechanism. It is probable that an immediate interim improvement could be made by eliminating corrections based on "tones" identified at frequencies below 500 Hz.

Finally, further experimental research is required into the perception of low frequency harmonic noise. Deficiencies in the scales were noted for application to helicopter noise and yet it has not been possible to relate these to any measured characteristics of the sounds themselves.

REFERENCES

1. McKennell, A. C., "Aircraft Noise Annlyance Around London (Heathrow) Airport," Wilson Committee on the Problem of Noise, Report SS 337, April 1963.

2. Stevens, S. S., "Assessment of Loudness Calculation Procedure, Mark VII," Harvard University, Draft, 1970.

3. Steinberg, J. C., "The Loudness of a Sound and Its Physical Stimulus," Physics Review, Vol. 26, p. 507, 1925.

4. Fletcher, H. and Munson, W. A., "Loudness, Its Definition, Measurement and Calculation," Journal of the Acoustical Society of America, Vol. 5, pp. 82-108, October 1933.

5. Fletcher, H., "Auditory Patterns," Review of Modern Physics, Vol. 12, pp. 47-65, 1940.

6. Stevens, S. S., "Calculation of the Loudness of a Complex Noise," Journal of the Acoustical Society of America, Vol. 28, No. 5, pp. 807-832, September 1956.

7. Zwicker, E., "Uber Psychologische und Methodische Grundlagen der Lautheit (Psychological and Methodical Foundations of Loudness)," Acustica, Vol. 3, pp. 237-258, 1958.

8. Kryter, K. D., "Possible Modifications to the Calculation of Perceived Noisiness," NASA Contractor Report CR-1636, August 1970.

9. Ollerhead, J. B., "Subjective Evaluation of General Aviation Aircraft Noise," FAA-NO-68-35, April 1968.

10. Adcock, B. D. and Ollerhead, J. B., "Effective Perceived Noise Level Evaluated for STOL and Other Aircraft Sounds," FAA-NO-70-5, May 1970.

11. Kryter, K. D., "Review of Research and Methods for Measuring the Loudness and Noisiness of Complex Sounds," NASA Contractor Report CR-422, April 1966.

12. U.S.A. Standard General Purpose Sound Level Meters, USAS 51.4, 1961.

13. Robinson, D. W. and Dadson, R. S., "A Redetermination of the Equal Loudness Relations for Pure Tones," British Journal of Applied Physics, Vol. 7, pp. 166-181, May 1956.

14. Robinson, D. W., "The Subjective Loudness Scale," Acustica, Vol. 7, pp. 217-233, 1957.

15. Kryter, K. D., and Pearsons, K. S., "Some Effects of Spectral Content and Duration on Perceived Noise Level," Journal of the Acoustical Society of America, Vol. 35, pp. 866-883, June 1963.

16. Stevens, S. S., "Concerning the Form of the Loudness Function," Journal of the Acoustical Society of America, Vol. 29, No. 5, pp. 603-606, May 1957.

17. Zwicker, E., Flottorp, G. and Stevens, S. S., "Critical Bandwidth in Loudness Summation," Journal of the Acoustical Society of America, Vol. 29, pp. 548-557, May 1957.

18. Zwicker, E., "Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen), " Journal of the Acoustical Society of America, Vol. 33, No. 2, p. 248, February 1961.

19. Zwicker, E., "Ein Graphisches Verfahren zur Bestimmung der Lautstarke und der Lautheit aus dem Terzpegeldiagram (Graphic Evaluation of Loudness from 1/3-Octave Band Sound Pressure Level Curves),"Frequenz, Vol. 13, No. 8, pp. 234-238, 1959.

20. Stevens, S. S., "The Measurement of Loudness," Journal of the Acoustical Society of America, Vol. 27, No. 5, pp. 815-829, September 1955.

21. U.S.A. Standard Procedure for the Computation of Loudness of Noise, USAS 53.4, 1968.

22. Stevens, S. S., "Procedure for Calculating Loudness: Mark VI," Journal of the Acoustical Society of America, Vol. 31, No. 11, pp. 1577-1585, November 1961.

23. Kryter, K. D., "Scaling Human Reactions to the Sound from Aircraft," Journal of the Acoustical Society of America, Vol. 31, No. 11, pp. 1415-1429, November 1959.

24. Kryter, D. D., "Concepts of Perceived Noisiness, Their Implementation and Application," Journal of the Acoustical Society of America, Vol. 43, No. 2, pp. 344-361, February 1968.

25. Little, J. W., "Human Response to Jet Engine Noises," Noise Control , 11-13, May/June 1961.

26. Robinson, D. W., "A Note on the Subjective Evaluation of Noise," Journal of Sound and Vibration, Vol. 1, No. 4, pp. 468-473, 1964.

27. Copeland, W. C. T., Davidson, I. M., Hargest, T. J., and Robinson, D. W., "A Controlled Experiment on the Subjective Effects of Jet Engine Noise," Journal of the Royal Aeronautical Society, Vol. 64, p. 33, 1960.

28. Robinson, D. W. and Bowsher, J. M., "A Subjective Experiment with Helicopter Noises," Journal of the Royal Aeronautical Society, Vol. 65, p. 635, 1961.

29. Kryter, K. D. and Pearsons, K. S., "Judged Noisiness of a Band of Random Noise Containing an Audible Pure Tone," Journal of the Acoustical Society of America, Vol. 38, pp. 106-112, 1965.

30. Pearsons, K. S., "The Effects of Duration and Background Noise Level on Perceived Noisiness," FAA Technical Report ADS-78, April 1966.

31. Little, J. W. and Mabry, J., "Sound Duration and Its Effect on Judged Annoyance," Paper presented at the 72nd Meeting of the Acoustical Society of America, Los Angeles, California, November 1966.

32. Pearsons, K. S., "Noisiness Judgements of Helicopter Flyovers," FAA Technical Report DS-67-1, January 1967.

33. Pearsons, K. S. and Horonjeff, R. D., "Category Scaling Judgement Tests on Motor Vehicle and Aircraft Noise," FAA Technical Report DS-67-8, July 1967.

34. Wells, R. J., "Recent Research Relative to Perceived Noise Level," Paper before 74th Meeting of the Acoustical Society of America, Miami, Florida, November 1967.

35. Mabry, J. E. and Little, J. W., "Laboratory Study of Jet Aircraft Noise Complaint Potential as Related to Context of the Noise," Paper presented at 74th Meeting of the Acoustical Society of America, Miami, Florida, November 1967.

36. Woodall, J. and Advisors, FAA Aircraft Development Service, Fourth Revised Draft of Proposed Noise Certification Criteria, June 1967.

37. Little, J. W. and Mabry, J. E., "Evaluation of Fifteen Scaling Units for Estimating 'Annoyance' of Jet Aircraft Flyovers," Paper before the 74th Meeting of the Acoustical Society of America, Miami, Florida, November 1967.

38. Hecker, M. H. L. and Kryter, K. D., "Comparisons Between Subjective Rating of Aircraft Noise and Various Objective Measures," FAA-NO-68-33, April 1968.

39. Pearsons, K. S., Horonjeff, R. D. and Bishop, D. E., "The Noisiness of Tones Plus Noise," NASA Contractor Report CR-1117, August 1968, (Also Journal of the Acoustical Society of America, Vol. 45, No. 3, pp. 742-750, 1969).

40. Hinterkeuser, E. G. and Sternfeld, H., "Subjective Response to Synthesized Flight Noise Signatures of Several Types of V/STOL Aircraft," NASA Contractor Report CR-1118, August 1968.

41. Wells, R. J., "Possible Modifications in the Computation of Perceived Noise Level," Paper before the 6th International Congress of Acoustics, Tokyo, August 1968.

42. Young, R. W. and Peterson, A., "On Estimating Noisiness of Aircraft Sounds," Journal of the Acoustical Society of America, Vol. 45, No. 4, pp. 834-838, 1969.

43. Pearsons, K. S. and Bennett, R. L., "The Effects of Temporal and Spectral Combinations on the Judged Noisiness of Aircraft Sounds," FAA NO-69-3, June 1969.

44. Federal Aviation Regulations, Part 21, Certification Procedure for Product and Parts; Part 36, Noise Standards; Aircraft Type Certification, December 1969.

45. Wells, R. J., "Calculation of Annoyance Level for Sounds Containing Multiple Pure Tones," Paper before 78th Meeting of the Acoustical Society of America, San Diego, California, November 1969.

46. Pearsons, K. S., and Wells, R. J., "Judged Noisiness of Sounds Containing Multiple Pure Tones," Paper before the 78th Meeting of the Acoustical Society of America, San Diego, California, November 1969.

47. Parry, H. J. and Stevens, J. K., "The Interpretation and Meaning of Laboratory Determinations of the Effect of Duration on the Judged Acceptability of Noise," Paper before the 78th Meeting of the Acoustical Society of America, San Diego, California, 1969.

48. Young, R. W., "Sound Level vs. Time for Noise Exposure," Paper before the 79th Meeting of the Acoustical Society of America, Atlantic City, April 1970.

49. Kryter, K. D., Johnson, P. J., and Young, J. R., "Judgement Tests of Flyover Noise from Various Aircraft," NASA Contractor Report CR-1635, August 1970.

50. Robinson, D. W., "The Subjective Basis for Aircraft Noise Limitation," Journal of the Royal Aeronautical Society, Vol. 71, pp. 396-400, June 1967.

51. Ollerhead, J. B., "The Noisiness of Diffuse Sound Fields at High Intensities," FAA-NO-70-3, August 1969.

52. Pinker, R. A., "Mathematical Formulation of the Noy Tables," Journal of Sound and Vibration, Vol. 8, pp. 488-492, 1968.

53. Dixon, W. J. and Massey, F. J., Introduction to Statistical Analysis, 3rd. Edition, McGraw-Hill Book Company, New York, N.Y., 1969.

APPENDIX A

FACILITY AND INSTRUMENTATION

A.1 Recording System

The instrumentation used for recording aircraft sounds in the field is shown in figure 27. Two alternative systems were used. The first of these utilized an a.c. tape recorder powered by batteries through an inverter. This system was cumbersome and was later replaced by a battery powered d.c. system. All measurements were made with a one-inch diameter B&K Model 4131 condenser microphone located approximately five feet from the ground with the plane of the diaphragm approximately parallel to the aircraft flight path. In the second system the same microphone was used with a gooseneck extension and windscreen. The two tape recorders are compatible with each other and have very similar frequency response characteristics. The system was initially calibrated using standards traceable to the National Bureau of Standards and checked periodically during recording sessions using a B&K Model 4220 pistonphone.

A.2 Paired-Comparison Tape-Making Procedure

The manner in which the sound pair arrangements were developed was described in section 3.2.3 of the main text. A total of 60 paired-comparison tapes were made, each comprising 25 pairs of sounds. The duration of the real aircraft sounds was variable, but the pairs were arranged with constant intervals of one second between two sounds of a pair, and six seconds between pairs.

A block flow diagram of the tape-making procedure and hardware is presented as figure 28. The twelve aircraft signals comprising one of the five groups of sounds were dubbed from the original field-recorded source tapes onto channels 1-12 of a CEC Model GR 2800 14-track one-inch FM tape recorder.

The simulated jet noise (Level 2 Reference) was copied onto channel 13, and the 1000 Hz octave-band of pink noise (Level 1 Standard) was written onto track 14. All of these signals were recorded at the same level.

As each signal was written, a 10 kHz tone was mixed with it for the required duration of the signal, in order to subsequently determine the beginning and end of the wanted portion of the sound. The one-inch tape was then spliced into a continuous loop, the length of which was dependent upon the duration of the longest aircraft signal being used. This loop was played back on a one-inch Sangamo Model 4700 FM recorder with a long loop capability, which acted as a data source for an automatic

compiler. This device, which was described in detail in reference 10, is a hardware package developed for this project which controls the entire tape-making sequence. The required aircraft and reference sounds are routed to the automatic compiler from the appropriate channels of the Sangamo 4700 tape-loop reproducer. A 50-position stepping-switch used in conjunction with a 50-connector patchboard determines both the input channel to be used and its required attenuation level, for each of the 50 sounds comprising the 25 sound pairs of any test tape.

Having selected one of the 50 positions, the system monitors the preselected channel as the tape loop goes round, until the superimposed 10 kHz control signal is detected to indicate the beginning of a sound record. The output tape recorder, a 1/4" direct record Ampex AG 500, is then turned on, and the aircraft sound (with the 10 kHz control signal removed by a notch filter) is written onto channel 1 of the test tape for as long as the control signal remains on. The data signal is ramped at the beginning and end of the sound (60 dB in 50 milliseconds), to minimize starting transients. When the 10 kHz control signal cuts off, the data signal is ramped down, and the tape recorder continues to run, without data input, to provide the intersound spacing (one second between sounds of a pair, and six seconds between pairs). The 10 kHz control signal is written onto channel 2 of the test tape simultaneously with the data, in order to provide a timing control during playback. A sequency relay then turns off the tape recorder, and energizes the main stepping switch to input the next sound through the patchboard.

The system incorporates intricate timers and lockouts, and has been found to operate faultlessly. Two patch panels are used, one for each tape of a pair, and the external patching between the compiler input and the tape-loop player output has only to be changed for each pair of test tapes.

Six of the final test tapes were made from each tape loop, the only hardware alterations necessary from one tape to the next being the changing of a pre-wired patchboard on the automatic compiler and modifications to the patching of the loop-player output to the compiler input. The tape-making system then functioned fully automatically, providing the required sound attenuation levels, order of signal presentation, and timing. About 45 minutes were required to make each final tape once the tape loop and patching arrangements were made ready.

Simultaneously with the making of the test tapes, and using the same procedures, master tapes of all 120 aircraft sounds were recorded, for subsequent one-third octave analysis (section 3.3.1).

A.3 Sound Replay System

The tests were conducted in a wide frequency range progressive wave chamber which was described in section 3.2.4 and illustrated in figure 17.

Under normal circumstances, the sound is generated at the end of a 36-foot long exponential horn which expands into a 1500 cubic foot room, more than half of which is filled with a set of 12-foot deep fiberglass wedges attached to the wall facing the horn. Up to five subjects are accommodated in the space between the wedges and the horn mouth, which is approximately 13 feet wide by 10 feet high. The walls of the room are constructed of 12-inch thick concrete which provide a high transmission loss to external sound.

In previous studies (ref. 57) it has been found that the best low frequency performance was obtained by baffle-mounting five speakers in a vertical array in the flare section of the horn, as shown in figure 17. The loudspeakers are 14.5-inch diameter JBL LE 15A's, and in this configuration can generate sound pressure levels in excess of 115 dB in the frequency range 20-5000 Hz, and above 118 dB between 25 and 4000 Hz.

At higher frequencies the problem is one of achieving uniform directional radiation. The large horn ceased to act as such at fairly low frequencies and the directional characteristics of the speaker dominated the sound pressure level variation across the working section. To minimize this variation, an Altec Model 805 B multicellular horn is used in combination with an Altec 855B driver, which is designed for operation at frequencies above 500 Hz.

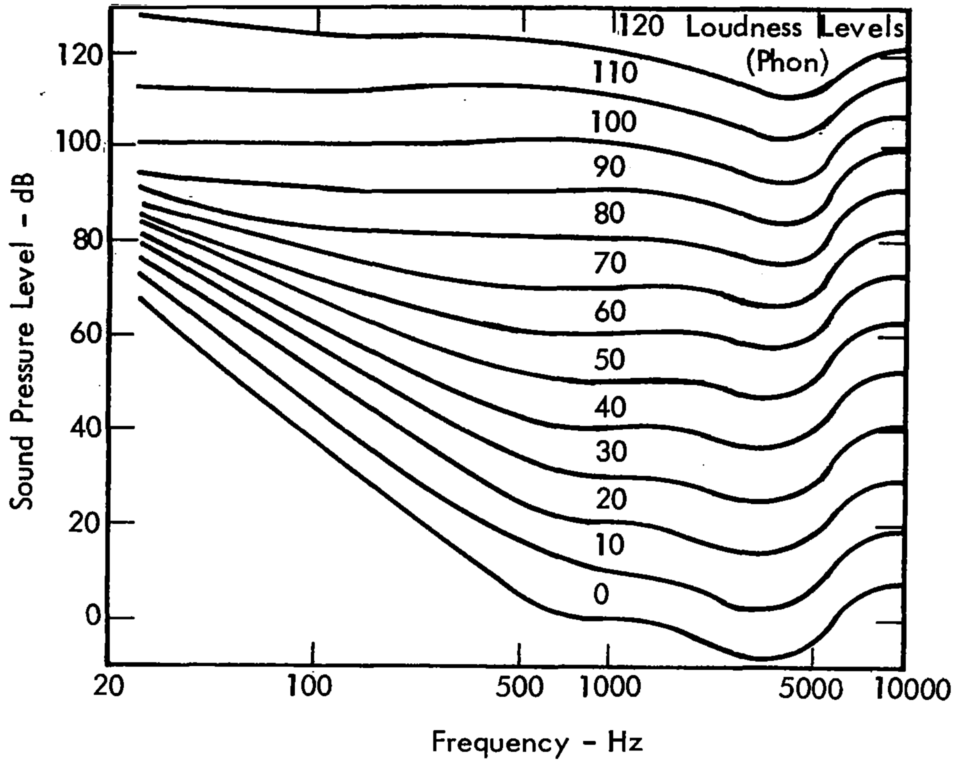
The complete instrumentation for the sound replay system is illustrated in figure 29. The signal from channel 1 of the tape recorder is first passed through a B&K Model 123, one-third octave band shaping filter in order to adjust the input spectrum to provide a flat response in the working section. On channel 2 a control tone is recorded which operates a ramp generator, the purpose of which is to suppress unwanted electronic noise in the chamber between sounds. This control tone additionally operates the sequence and pair number indicator lights, which are placed in front of the subjects. The low frequency portion of the signal (below 600 Hz) is filtered and passed to a 300 watt solid state amplifier which drives the five low frequency speakers. The high frequency horn is driven by a similar amplifier. The sound at the seating positions was at all times monitored by a one-half inch diameter B&K Model 4133 condenser microphone connected to an audio frequency spectrometer and graphic level recorder. The microphone was located centrally between the two middle seats, at which position the overall system frequency response was adjusted to maintain a flat response. This was done prior to each test by inserting a "pink" noise signal at the shaping filter input and adjusting for a flat one-third octave band spectrum at the microphone.

APPENDIX B

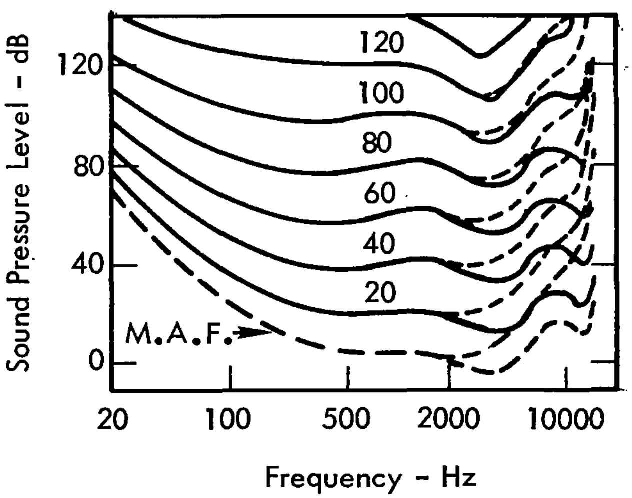
SPECTRA AND TIME HISTORIES OF THE 120 AIRCRAFT SOUNDS

It is not possible to include the complete set of 1/3-octave analyses in this report, but to give some indication of the characteristics of the 120 aircraft sounds studied, two plots are presented in figure 30 for each sound. The first of these is the 1/3-octave band spectrum corresponding to the first instant when the overall sound pressure level reaches its maximum value for the flyover. The second is the time history of the overall sound pressure level, plotted at 0.5-second intervals, during the two 7.5 second periods preceding and following the time instant ($t = 0$) corresponding to the 1/3-octave spectrum.

An attempt has been made to group the sounds by category so that the order thus differs from those in Tables II and V. It should also be noted that the absolute levels correspond to those in Table V, i.e. the levels at which the sounds were played to the subjects rather than the levels at which the sounds were recorded (Table II).

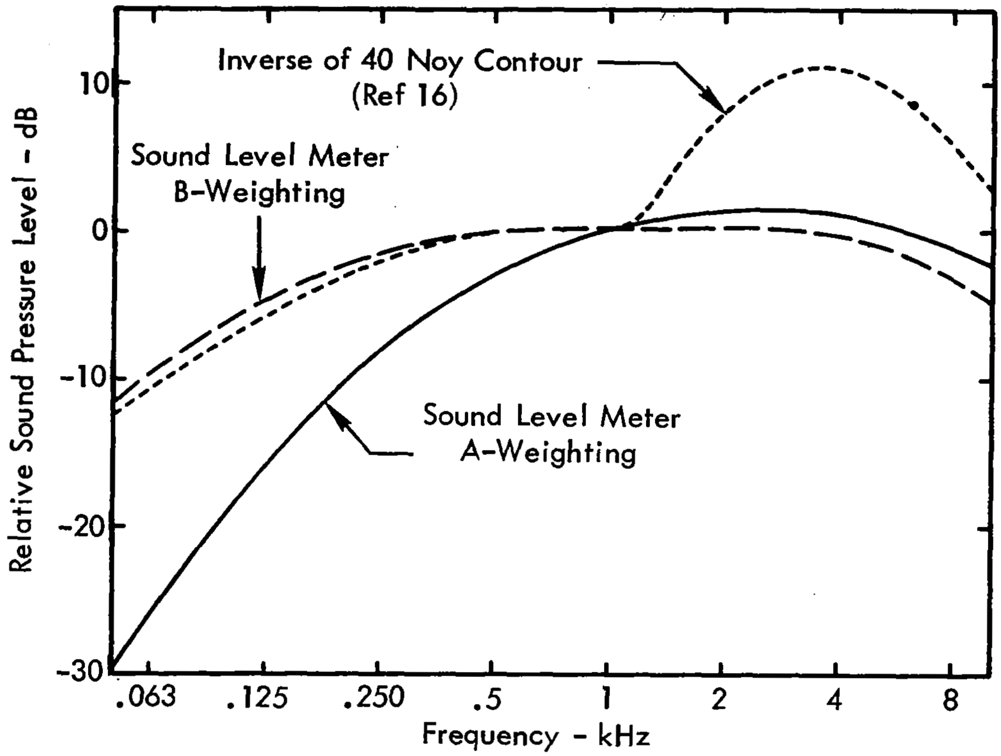


(a) After Fletcher and Munson (Ref 4), 1933



(b) After Robinson and Dadson (Ref 15), 1956

Figure 1: Equal Loudness Contours for Pure Tones



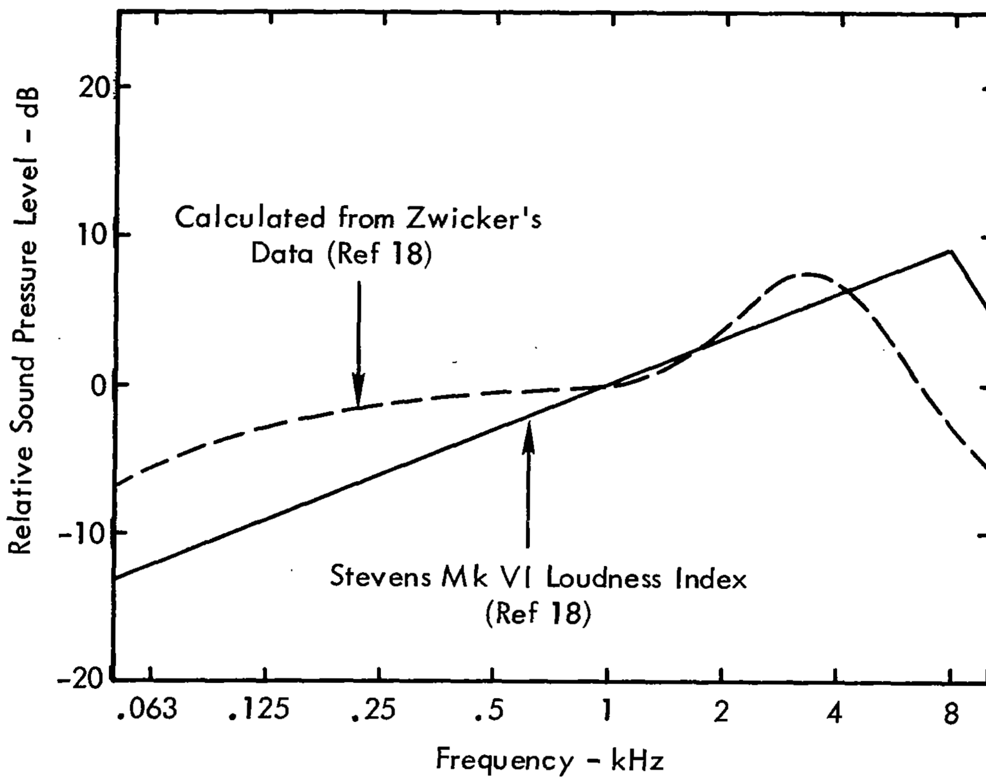
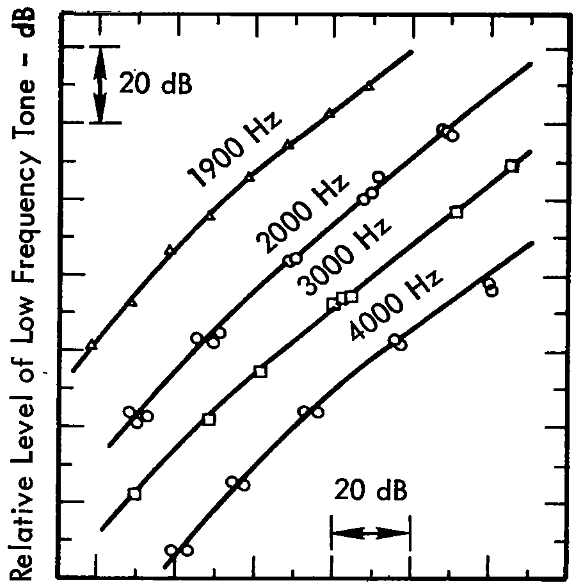


Figure 2. Some Frequency Weighting Functions
Based on Equal Magnitude Contours



Relative Level of Higher
Frequency Tone - dB

(Curves derived from loudness matches between a reference tone and a higher frequency comparison tone. Reference is at 700 Hz (triangles) or 1000 Hz (dots and squares.)

Figure 3. Variation of Loudness Growth with Frequency (After Stevens, Ref. 2).

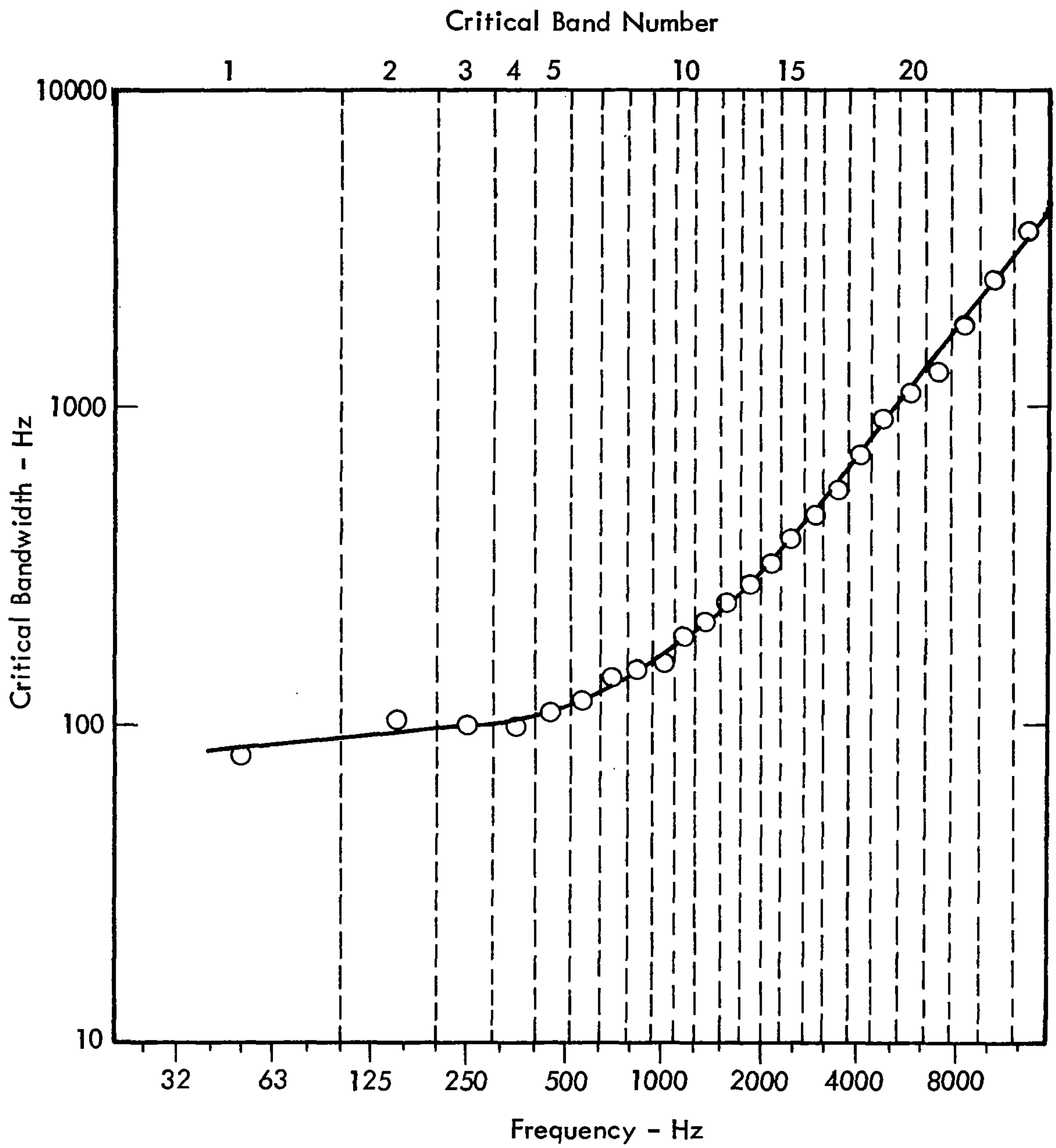


Figure 4. Division of the Audible Frequency Range into Critical Bands (After Zwicker - Ref. 18)

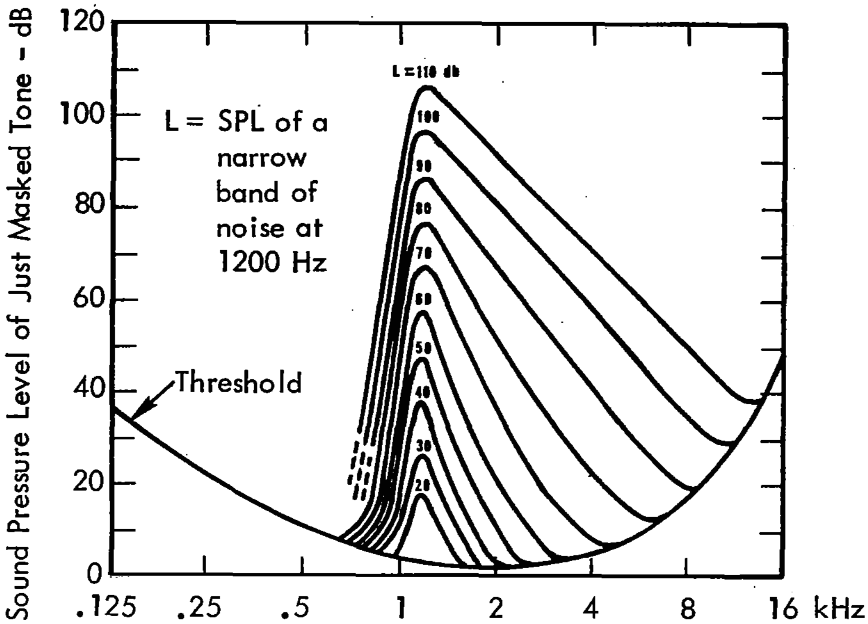


Figure 5. Masking of Tones by a Narrow Band of Noise of Sound Pressure L (from Zwicker - Ref. 7).

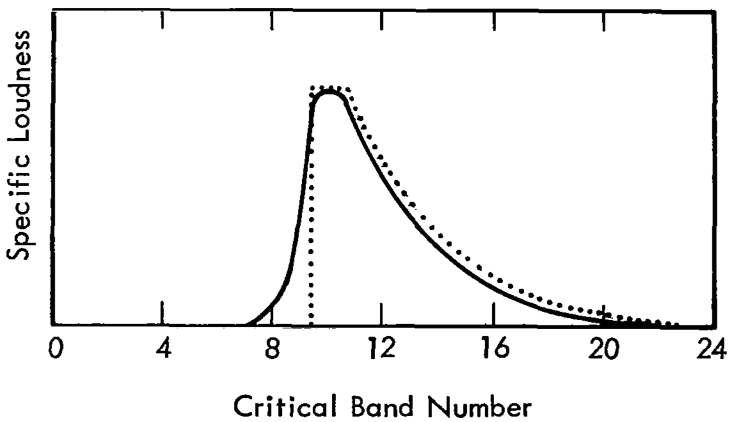


Figure 6. Calculated Specific Loudness of a 1200 Hz Tone at 100 dB (From Ref. 7).

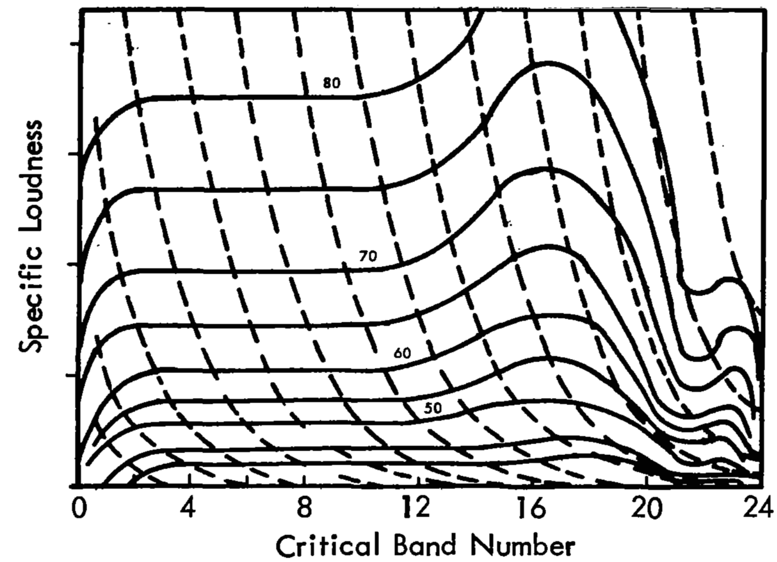


Figure 7. Zwicker's Original Loudness Calculation Chart (Reference 7)

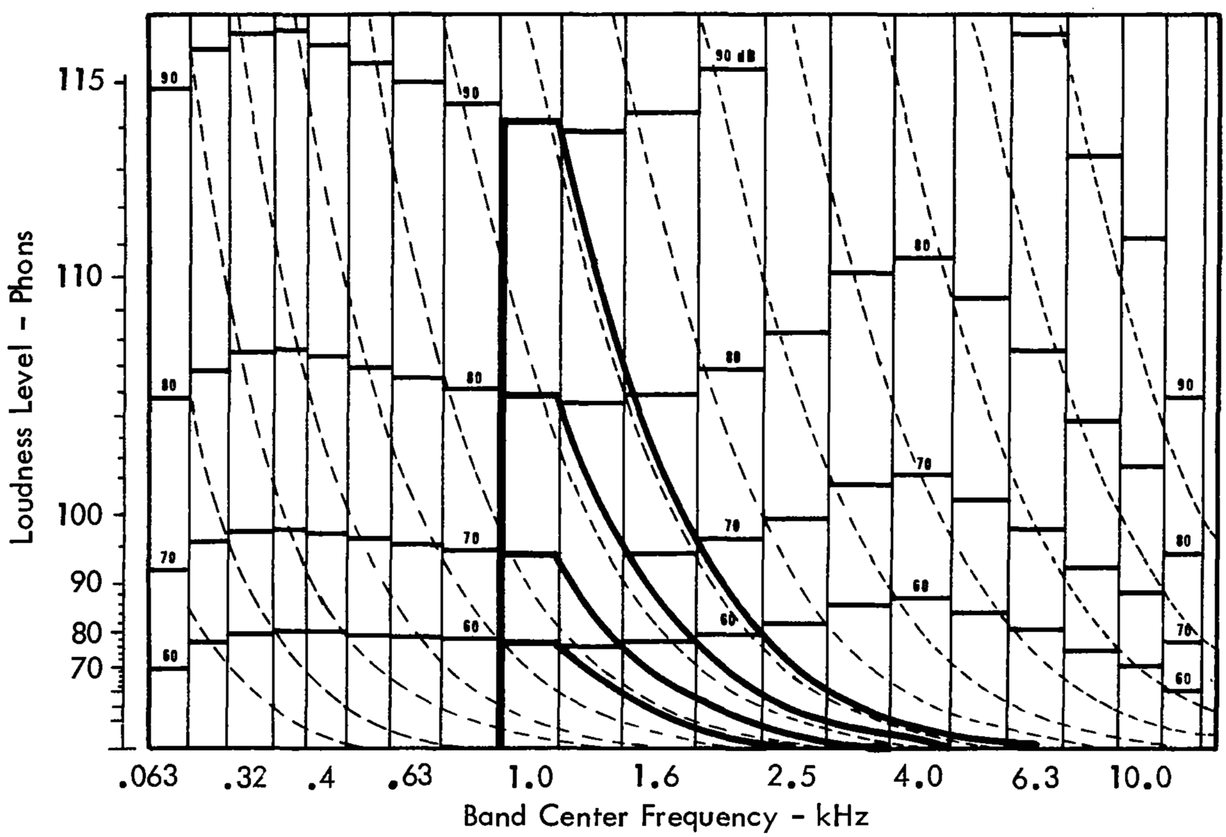


Figure 8. Loudness Calculation Chart for 1/3 - Octave Band Spectra With Demonstrated Use for 1/3 - Octave Bands of Noise Centered at 1000 Hz (from Zwicker - Reference 19)

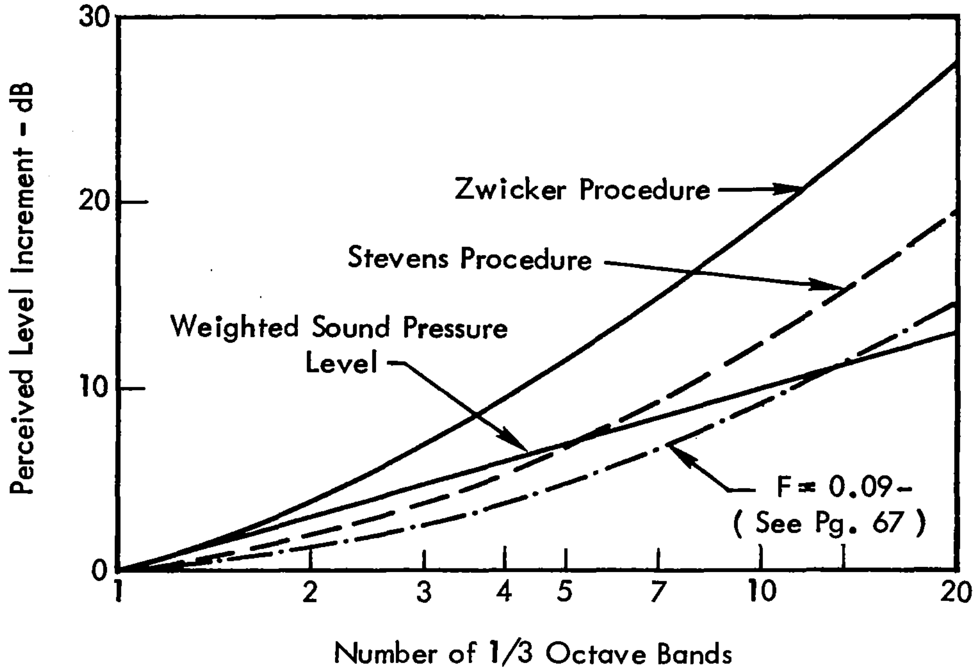


Figure 9. Bandwidth Summation of Perceived Level by Three Different Procedures

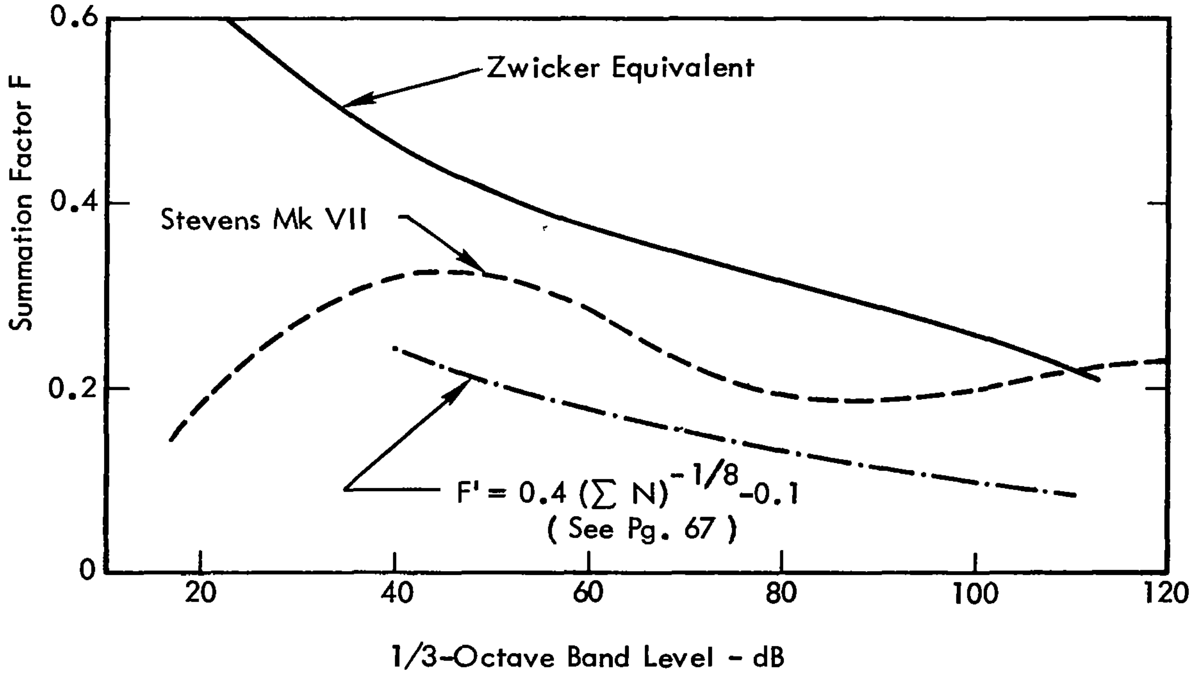


Figure 10. Variation of the Summation Factor F with Sound Pressure Level.

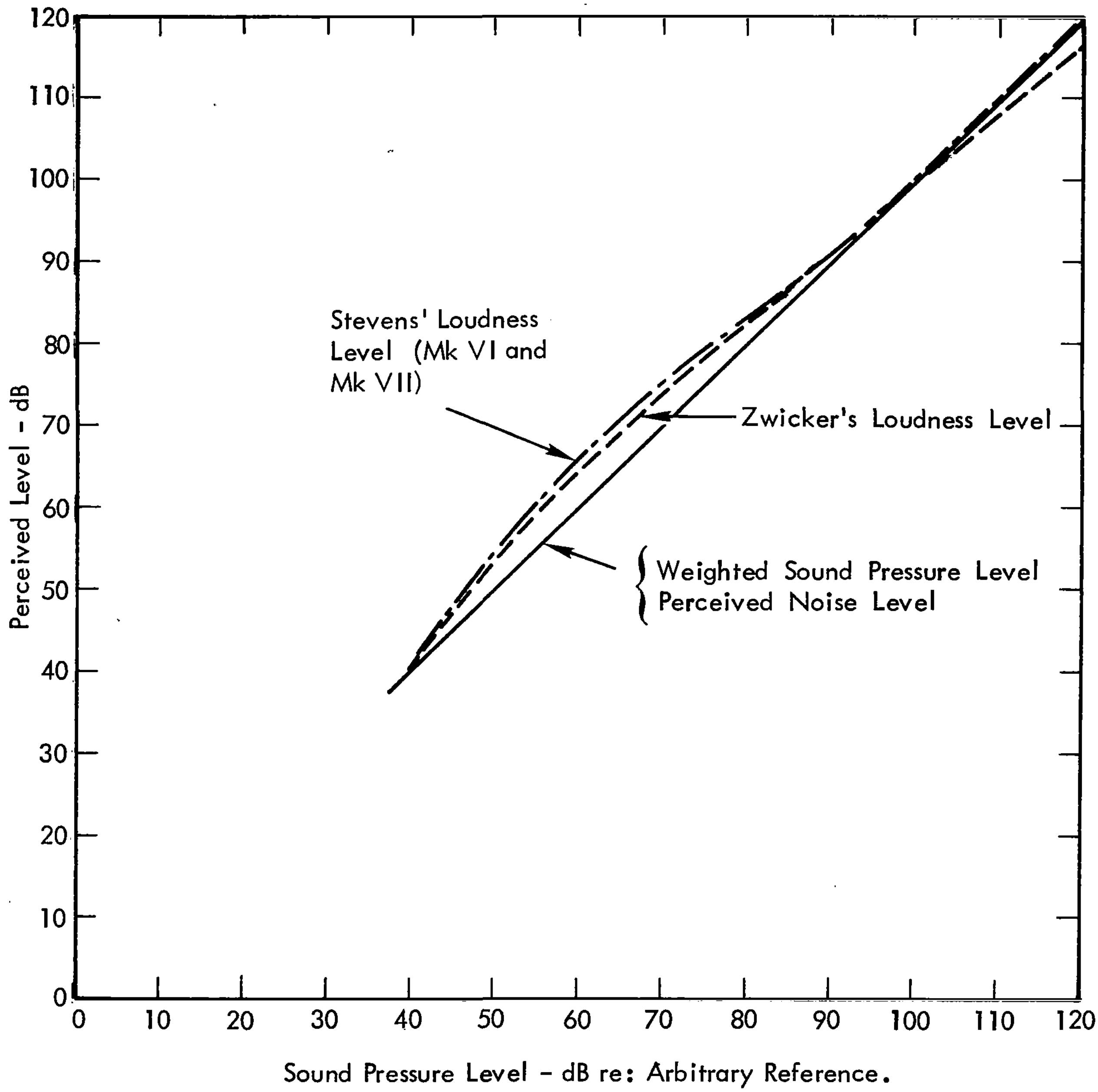


Figure 11. A Comparison of Perceived Level Growth Functions.

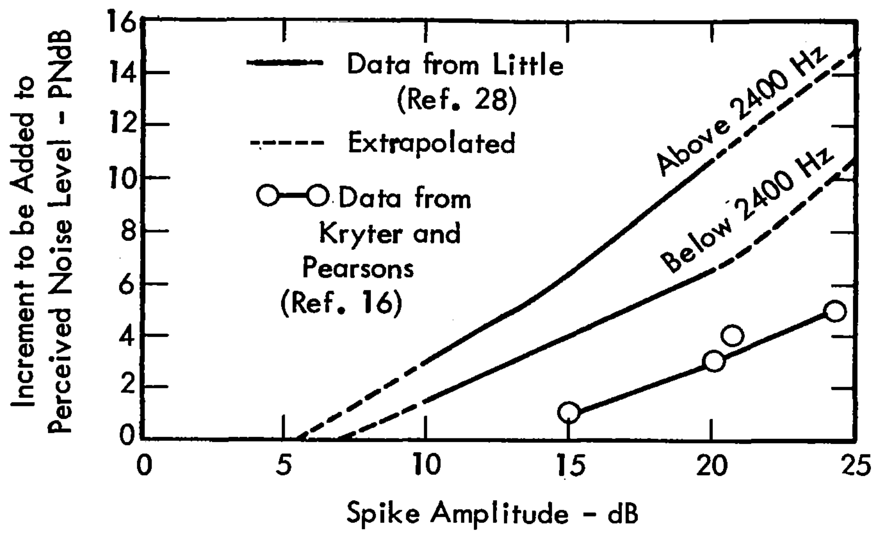
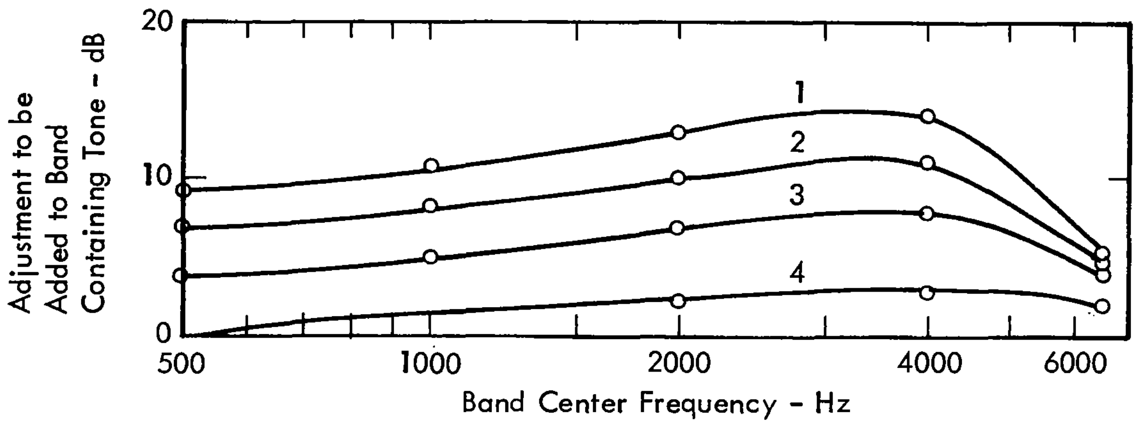


Figure 12. Tone Correction to PNL as a Function of Spike Amplitude Above Background level in 1/24th Octave Bands.



Curve Number	1	2	3	4
Tone Level - Noise Level*	25	15	5	-5
(Tone & Noise) Level - Level of Adjacent Band*	25	15	6	1

*Measured in 1/3 Octave Bands

Figure 13. Kryter-Pearsons Tone Correction.

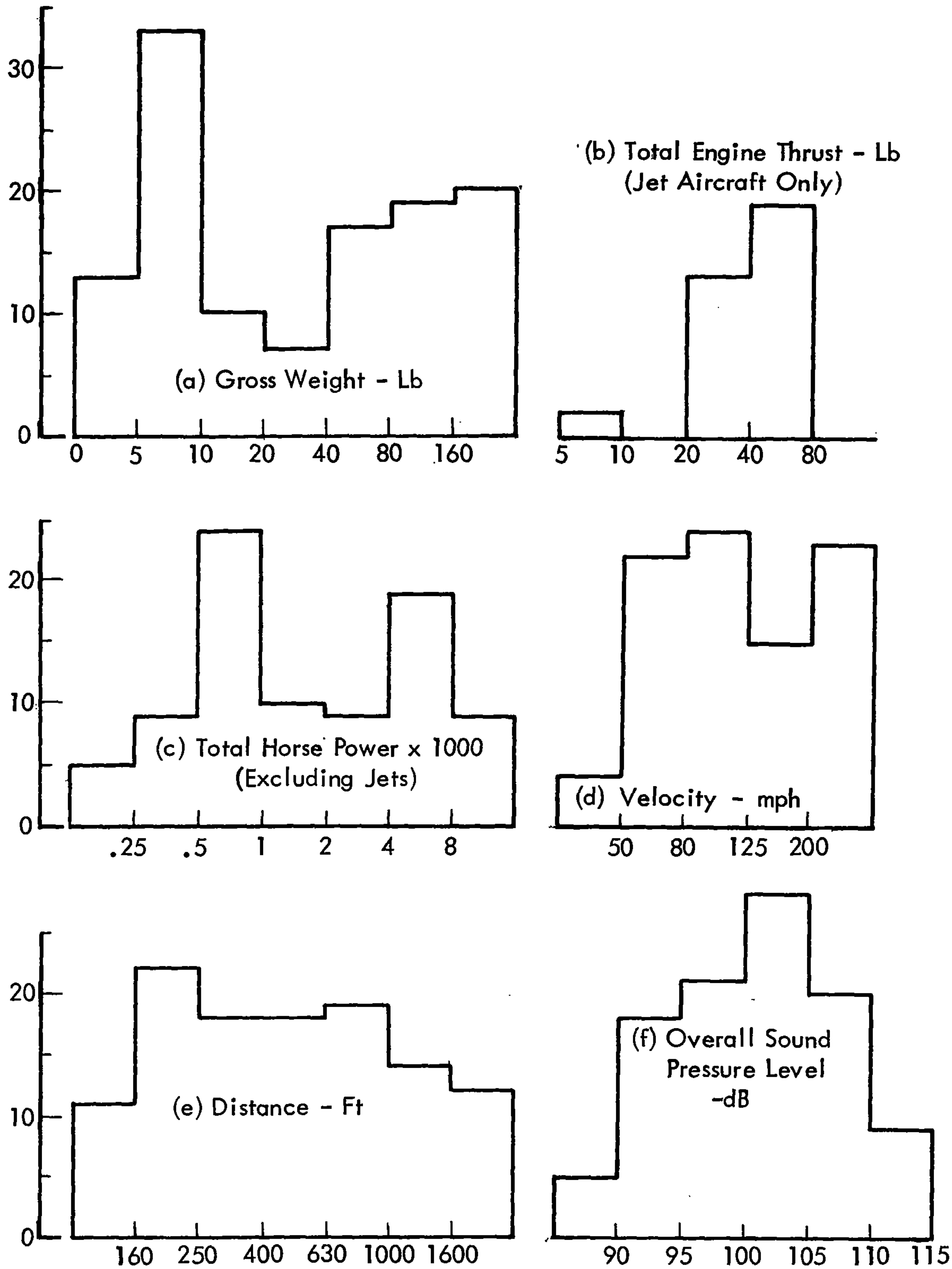


Figure 14. Histograms of Estimated Aircraft Data.

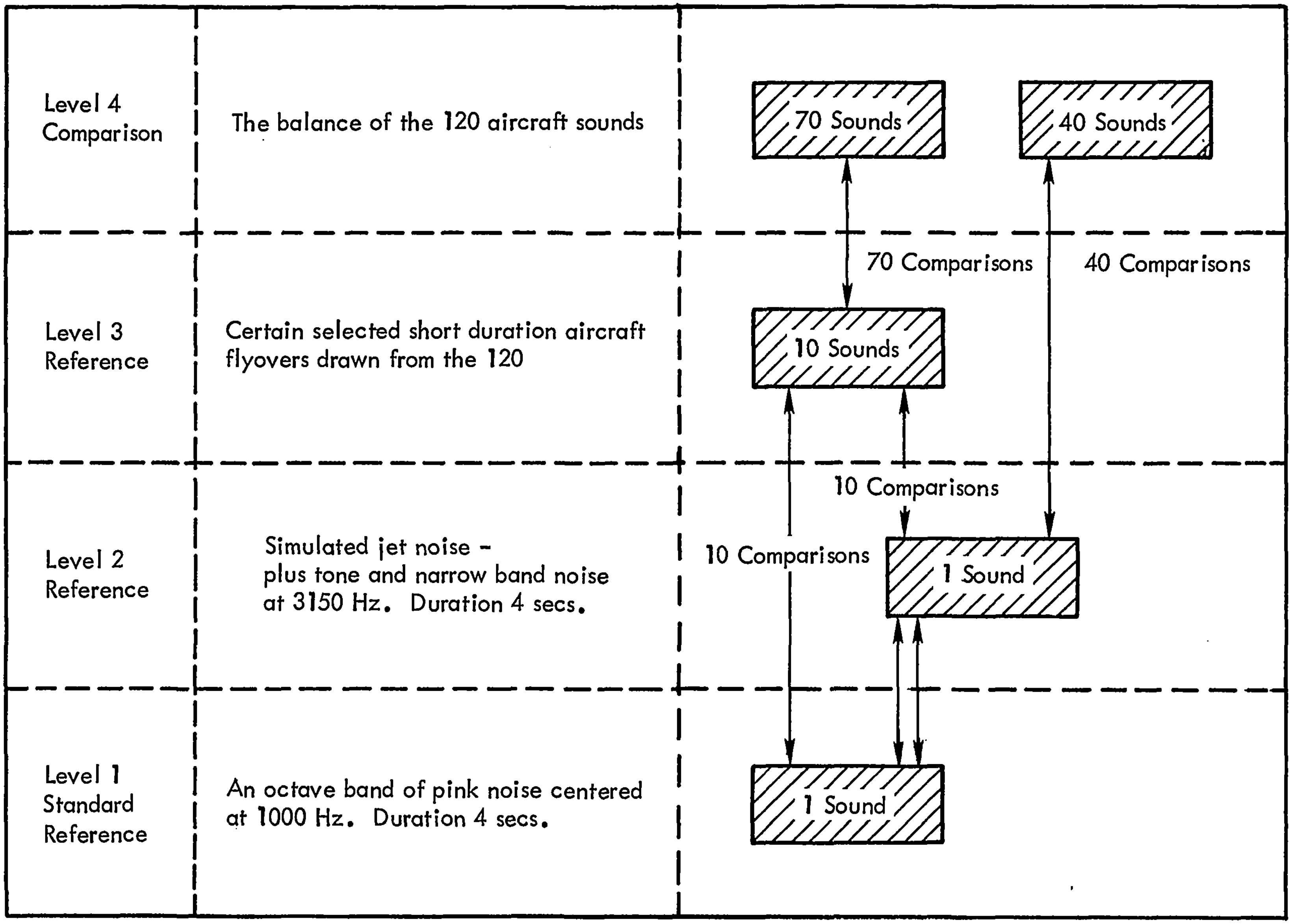
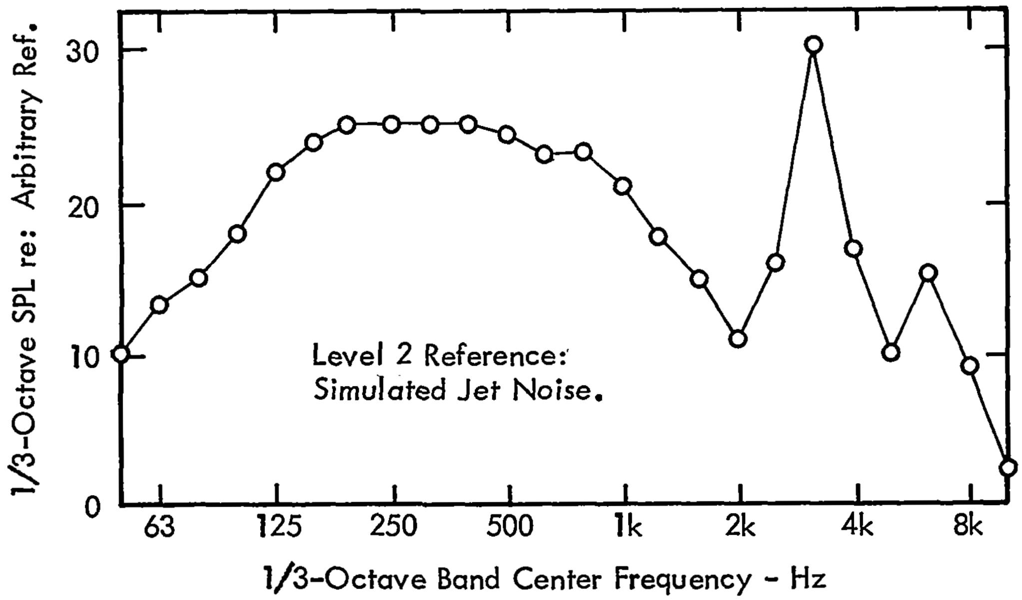


Figure 15. The Basic Sound Comparison Arrangement.



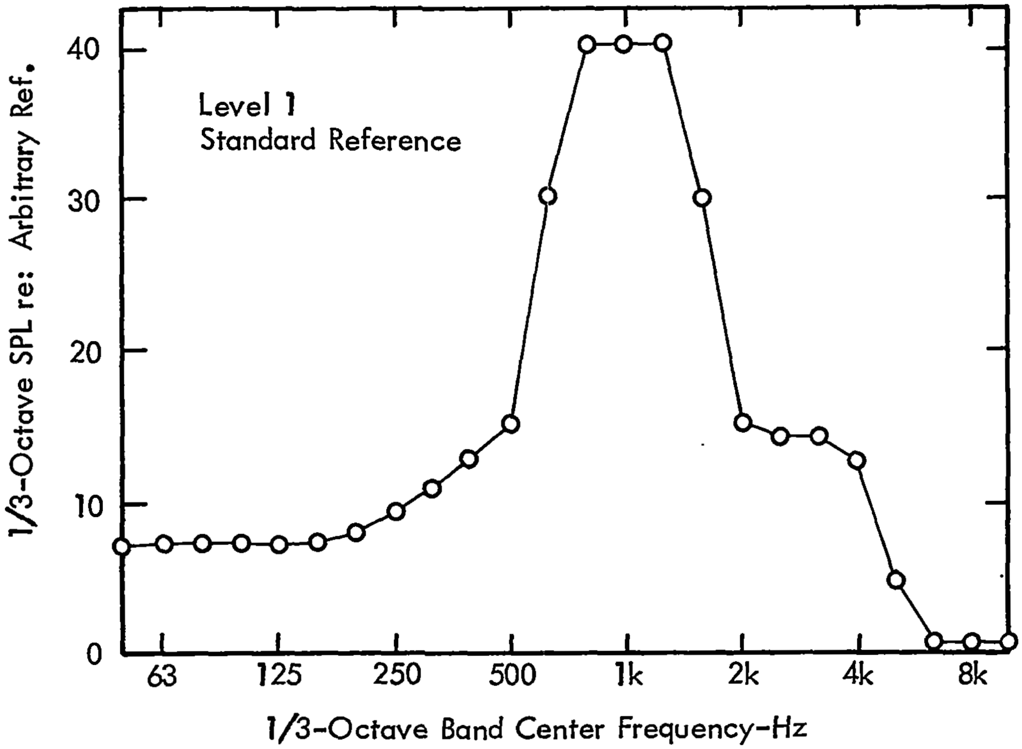


Figure 16. 1/3-Octave Band Spectra of Reference Sounds.

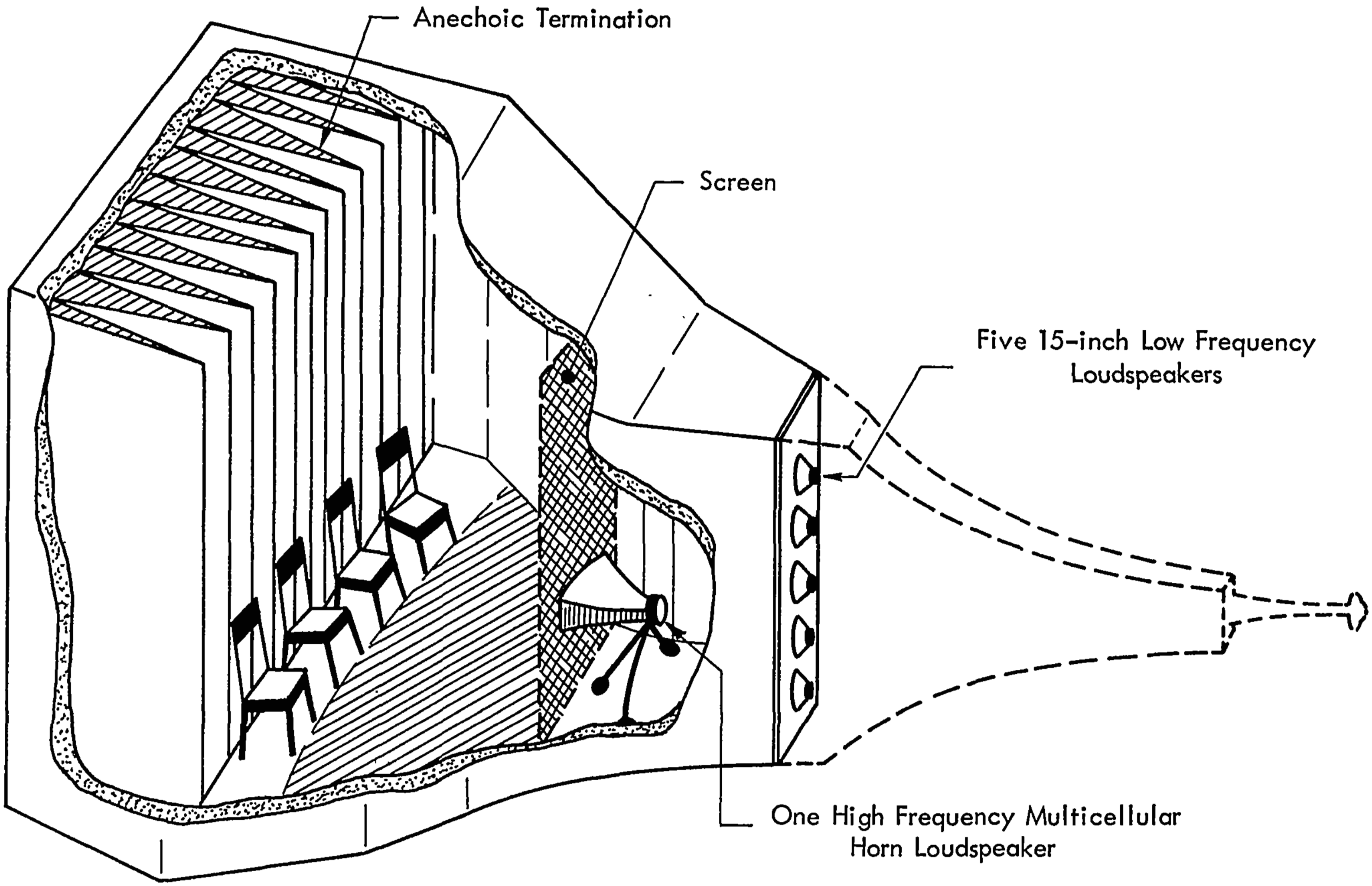


Figure 17. Cutaway Drawing of Progressive Wave Facility

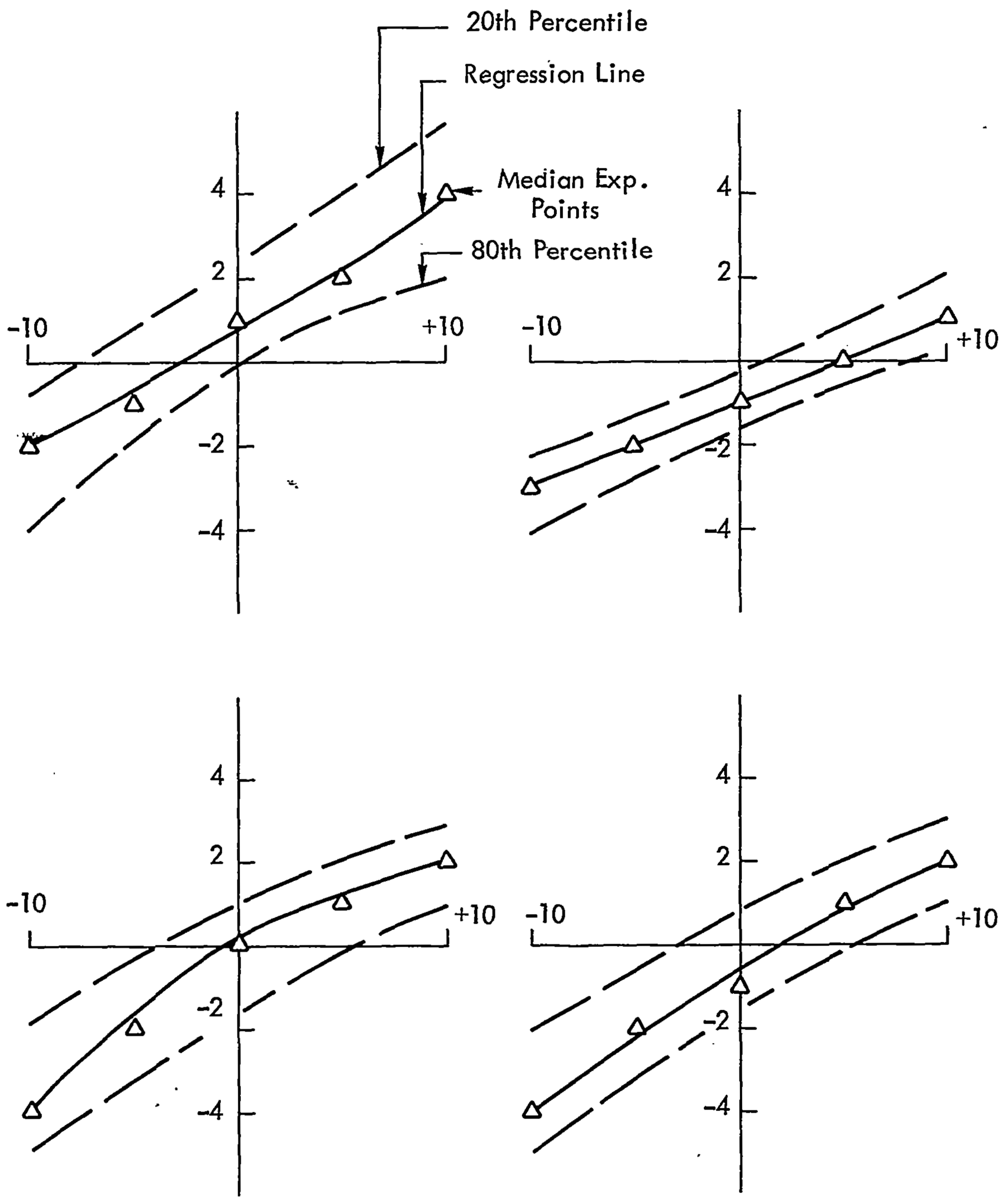
SOUND PAIR SET 7

SC.	F.		SC.	F.		SC.	F.		SC.	F.		SC.	F.	
-10	0		-10	0		-10	0		-10	0		-10	0	
-9	0		-9	0		-9	0		-9	0		-9	0	
-8	0		-8	0		-8	0		-8	0		-8	0	
-7	0		-7	0		-7	0		-7	0		-7	0	
-6	0		-6	0		-6	0		-6	0		-6	0	
-5	0		-5	0		-5	0		-5	0		-5	0	
-4	3	***	-4	0		-4	0		-4	0		-4	0	
-3	6	******	-3	2	**	-3	0		-3	0		-3	0	
-2	13	*************	-2	8	********	-2	1	*	-2	0		-2	1	*
-1	6	******	-1	10	**********	-1	1	*	-1	1	*	-1	1	*
0	3	***	0	7	*******	0	16	****************	0	5	*****	0	0	
1	1	*	1	4	****	1	11	***********	1	11	***********	1	3	***
2	0		2	1	*	2	2	**	2	10	**********	2	15	***************
3	0		3	0		3	1	*	3	5	*****	3	7	*******
4	0		4	0		4	0		4	0		4	3	***
5	0		5	0		5	0		5	0		5	2	**
6	0		6	0		6	0		6	0		6	0	
7	0		7	0		7	0		7	0		7	0	
8	0		8	0		8	0		8	0		8	0	
9	0		9	0		9	0		9	0		9	0	
10	0		10	0		10	0		10	0		10	0	

SOUND PAIR SET 8

SC.	F.		SC.	F.		SC.	F.		SC.	F.		SC.	F.	
-10	0		-10	0		-10	0		-10	0		-10	0	
-9	0		-9	0		-9	0		-9	0		-9	0	
-8	0		-8	0		-8	0		-8	0		-8	0	
-7	0		-7	0		-7	0		-7	0		-7	0	
-6	0		-6	0		-6	0		-6	0		-6	0	
-5	0		-5	0		-5	0		-5	0		-5	0	
-4	0		-4	0		-4	0		-4	0		-4	0	
-3	4	****	-3	0		-3	0		-3	1	*	-3	0	
-2	13	*************	-2	8	********	-2	1	*	-2	1	*	-2	1	*
-1	12	************	-1	11	***********	-1	1	*	-1	1	*	-1	0	
0	2	**	0	6	******	0	12	************	0	4	****	0	4	****
1	1	*	1	6	******	1	8	********	1	6	******	1	5	*****
2	0		2	1	*	2	8	********	2	10	**********	2	9	*********
3	0		3	0		3	2	**	3	7	*******	3	11	***********
4	0		4	0		4	0		4	2	**	4	2	**
5	0		5	0		5	0		5	0		5	0	
6	0		6	0		6	0		6	0		6	0	
7	0		7	0		7	0		7	0		7	0	
8	0		8	0		8	0		8	0		8	0	
9	0		9	0		9	0		9	0		9	0	
10	0		10	0		10	0		10	0		10	0	

Figure 18. Histogram of Subjective Response Scores



Abscissae: Level of Reference Sound with Respect to its Nominal Playback Level, dB

Ordinates: Subjective Responses with Bias Subtracted out, in Arbitrary Units

Figure 19. Determination of Subjectively Judged Levels; Some Typical Examples.

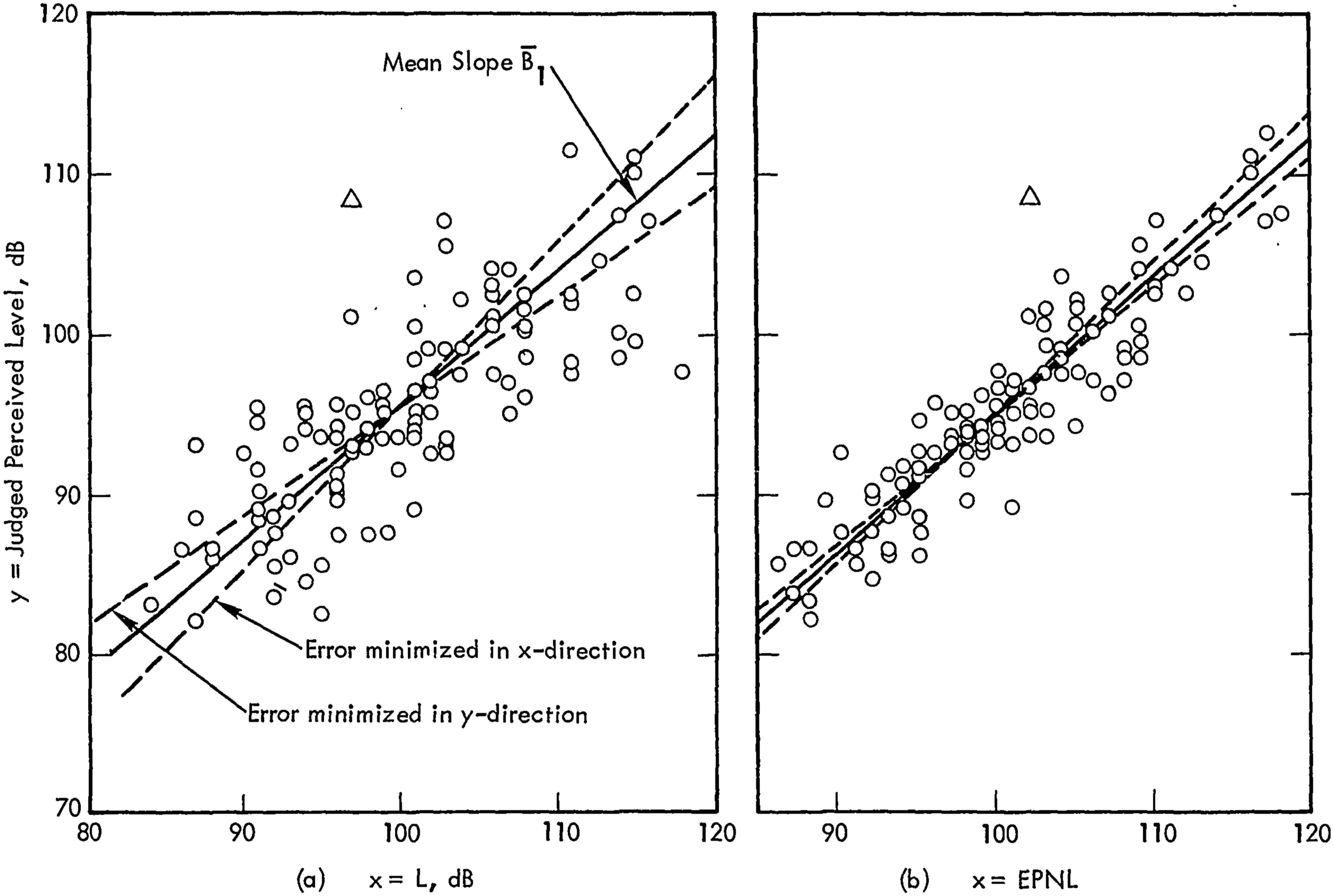


Figure 20. Scatter Diagrams, Judged v. Calculated Levels, All Sounds

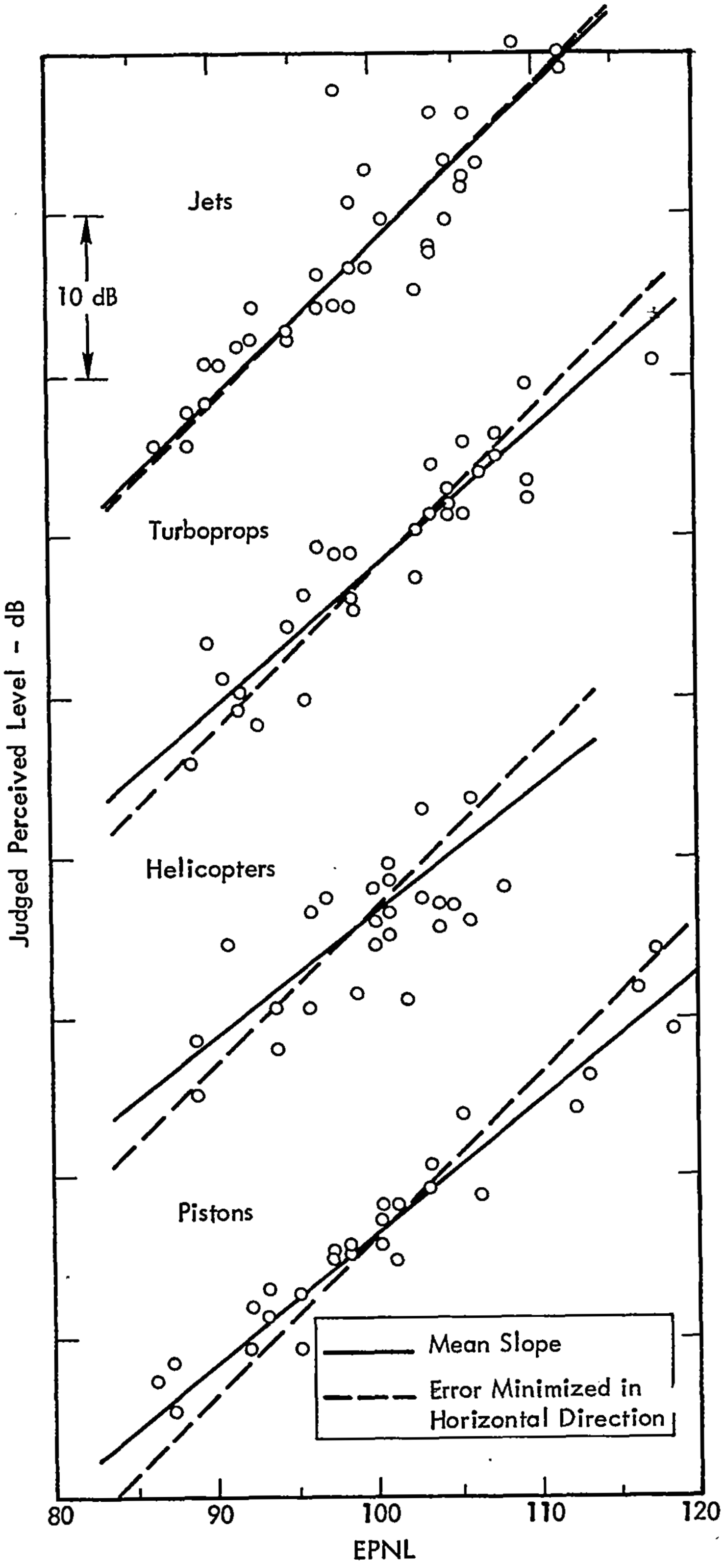
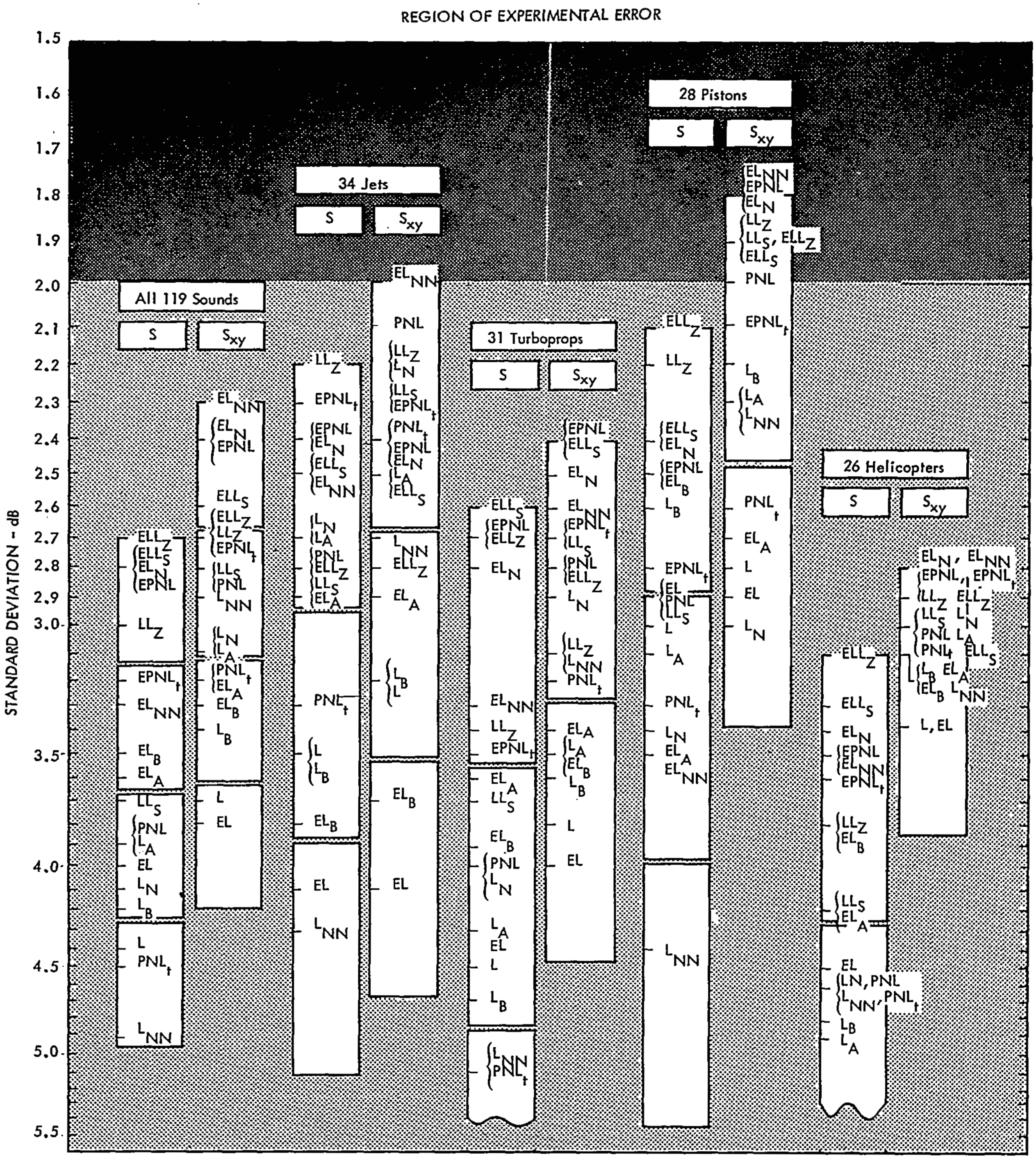


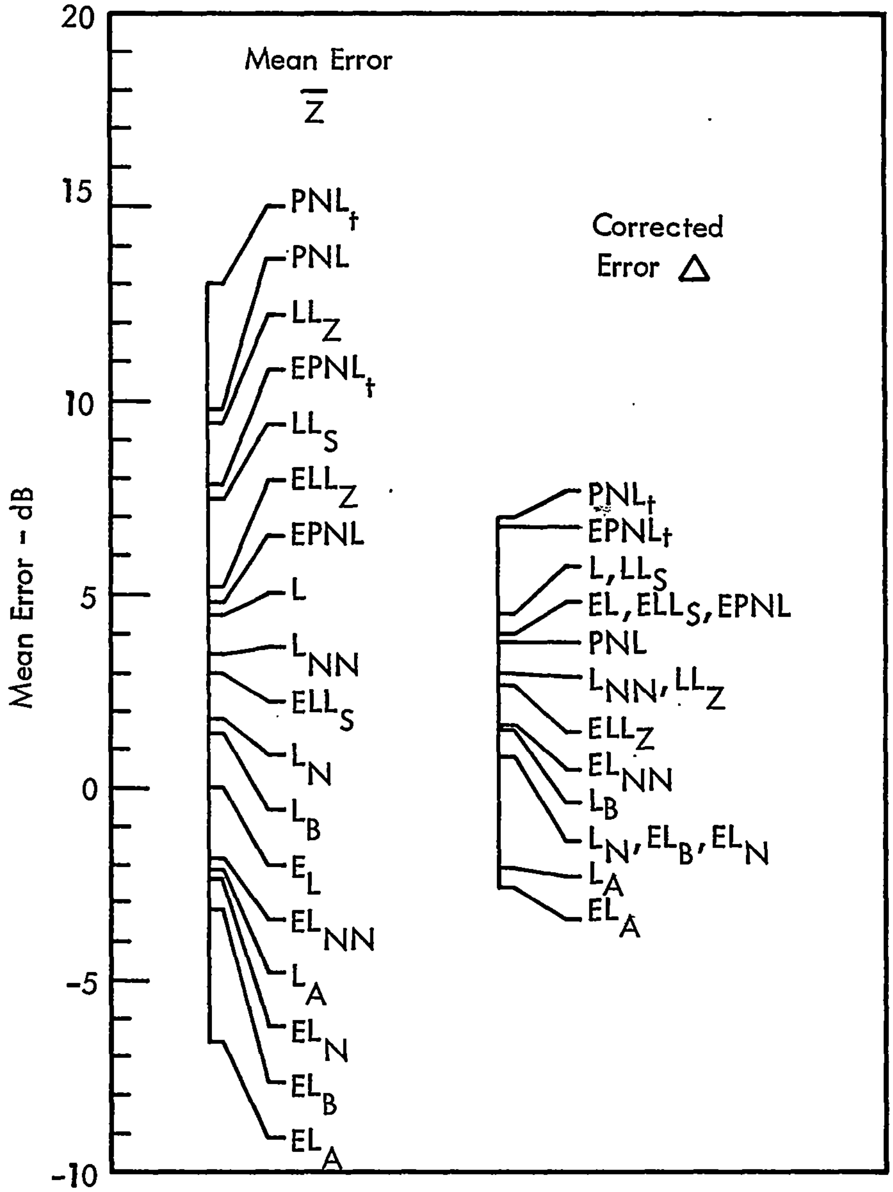
Figure 21. Scatter Diagrams for Sounds in Different Aircraft Categories.



S = standard deviation off difference: (calculated level) - (judged level)

S_{xy} = standard error of estimate (standard deviation about regression line $y = B_0 + B_1x$)

Figure 22. Consistency of Rating Scales When Applied to Sounds of Different Aircraft Categories.



$\overline{Z}$ = Average value of (Calculated Perceived Level of Aircraft Sound) - (Overall Level of Reference)

△= Average Value of (Calculated Perceived Level of Aircraft Sound) - (Calculated Perceived Level of Reference)

Figure 23. Accuracy of Scales for 119 Sounds.

(a) Jets

(b) Turboprops and Pistons

(c) Helicopters

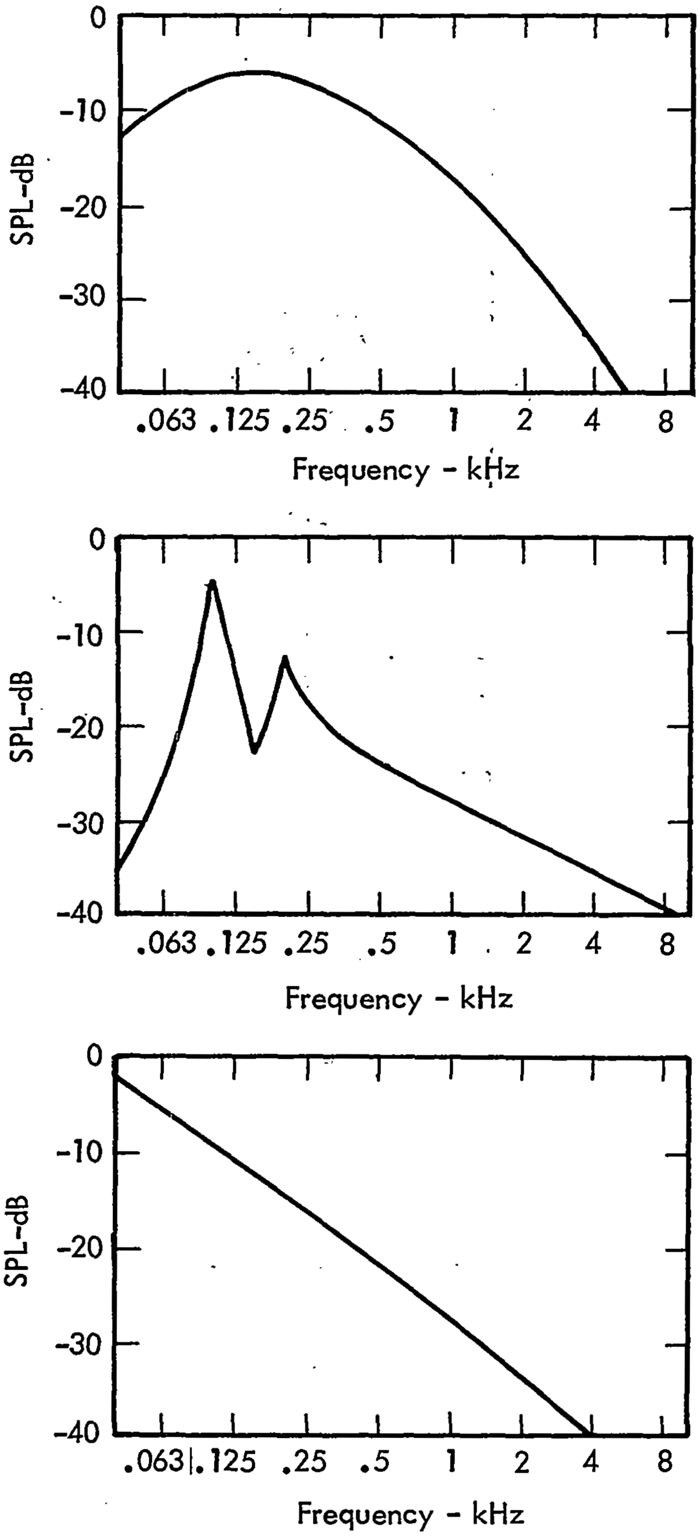


Figure 24. "Typical" 1/3 Octave Band Level Spectra for Different Aircraft Category Sounds.

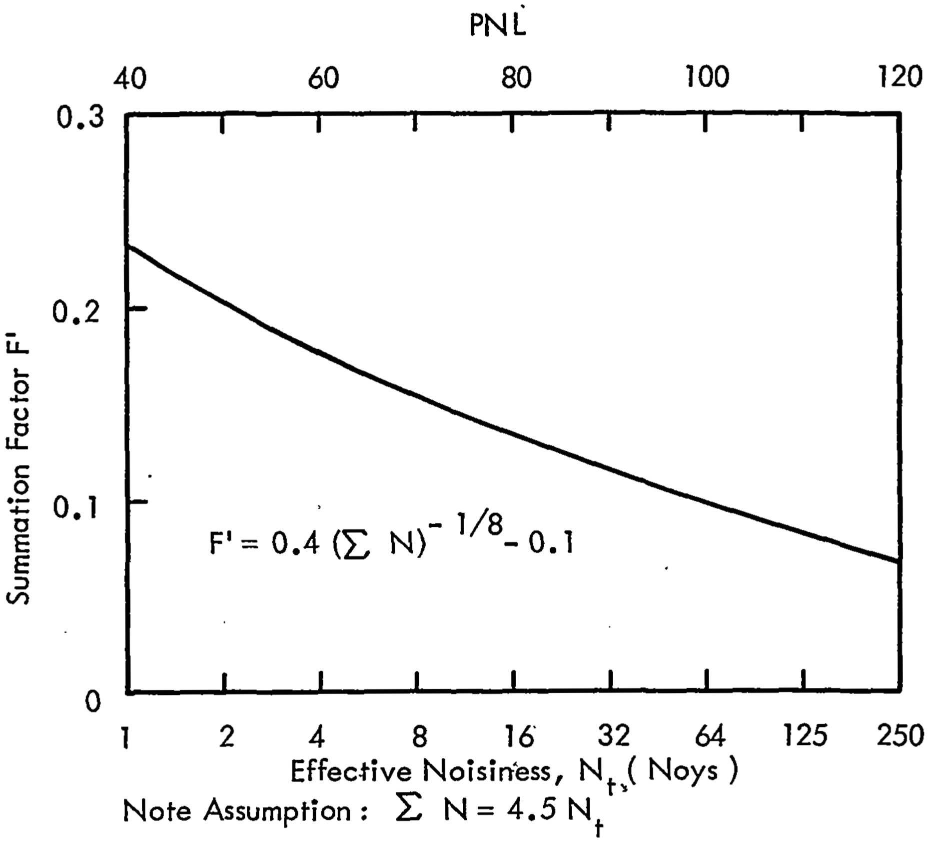


Figure 25. Revised Summation Factor F'.

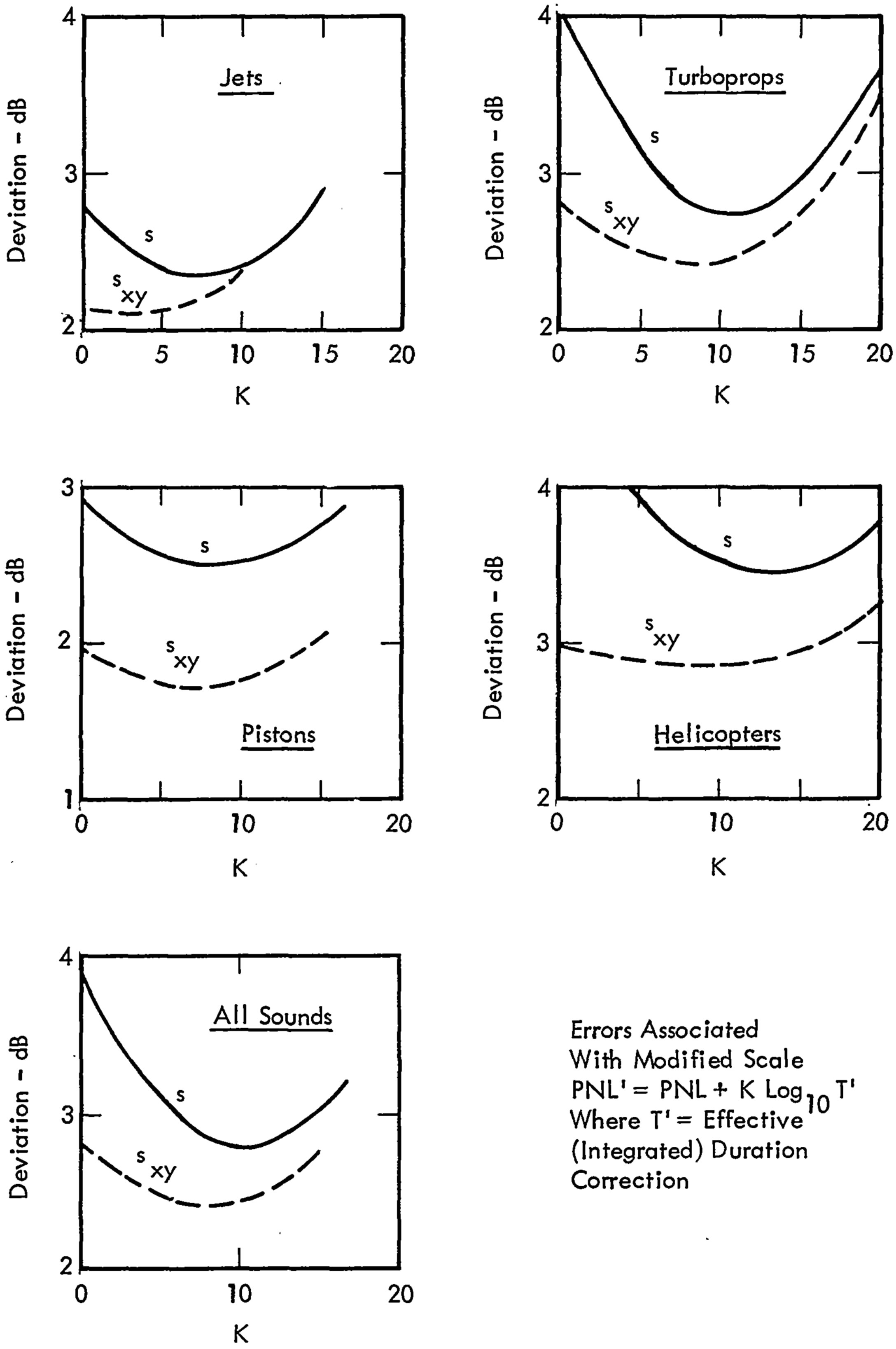


Figure 26. Optimum Duration Corrections.

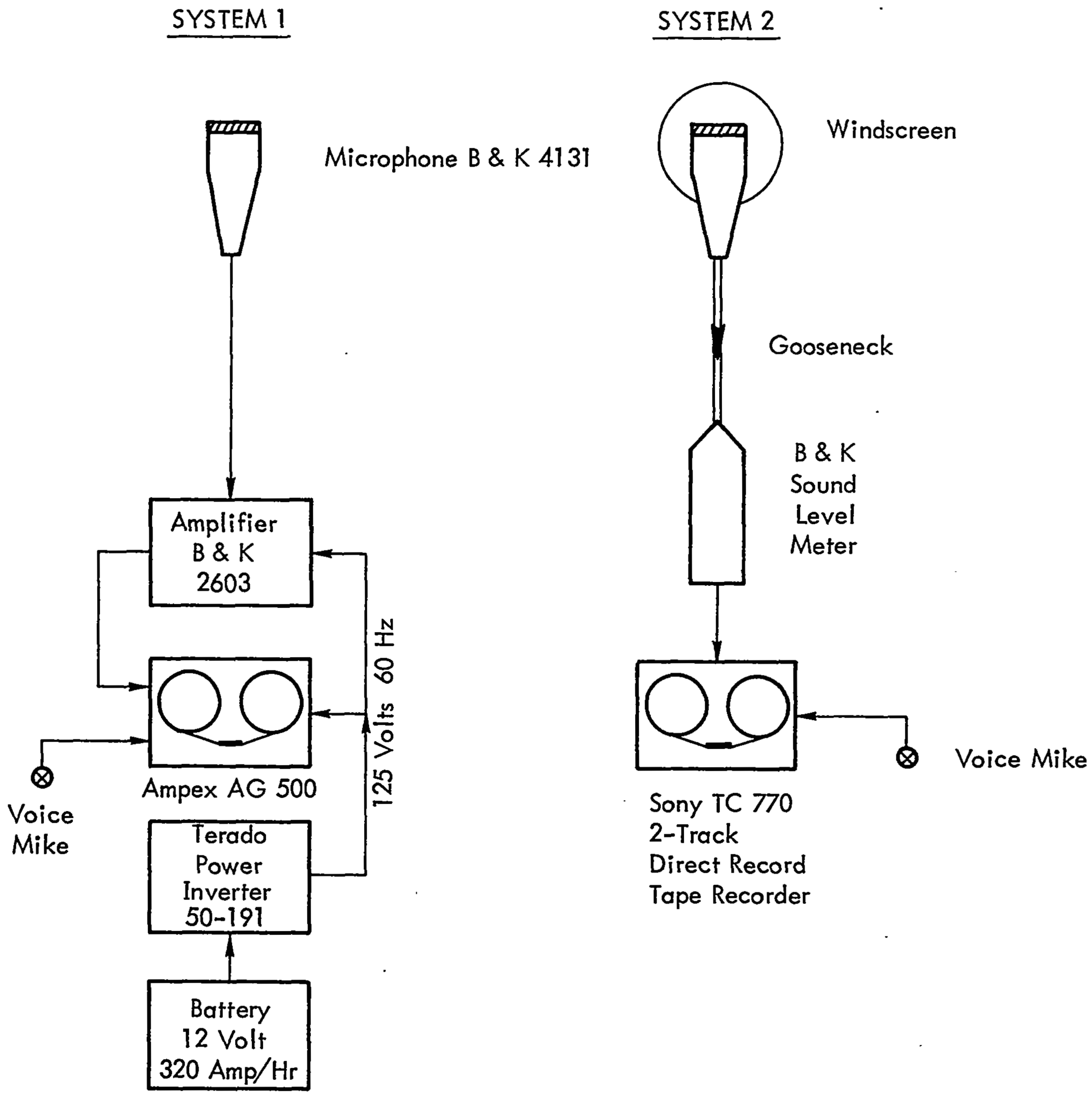


Figure 27. Instrumentation Used for Recording Aircraft Noise

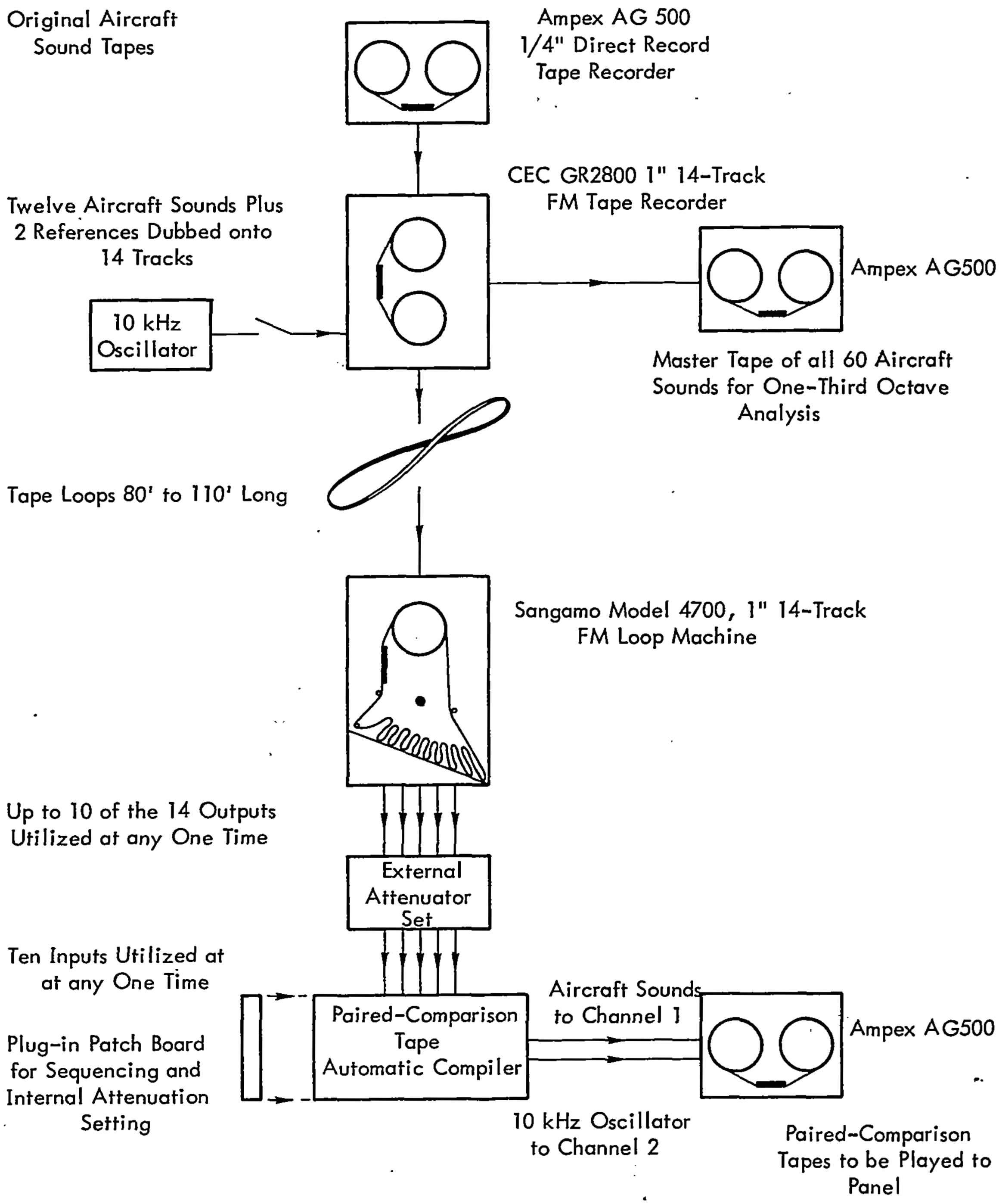


Figure 28. Schematic of Tape-Making Procedure and Hardware

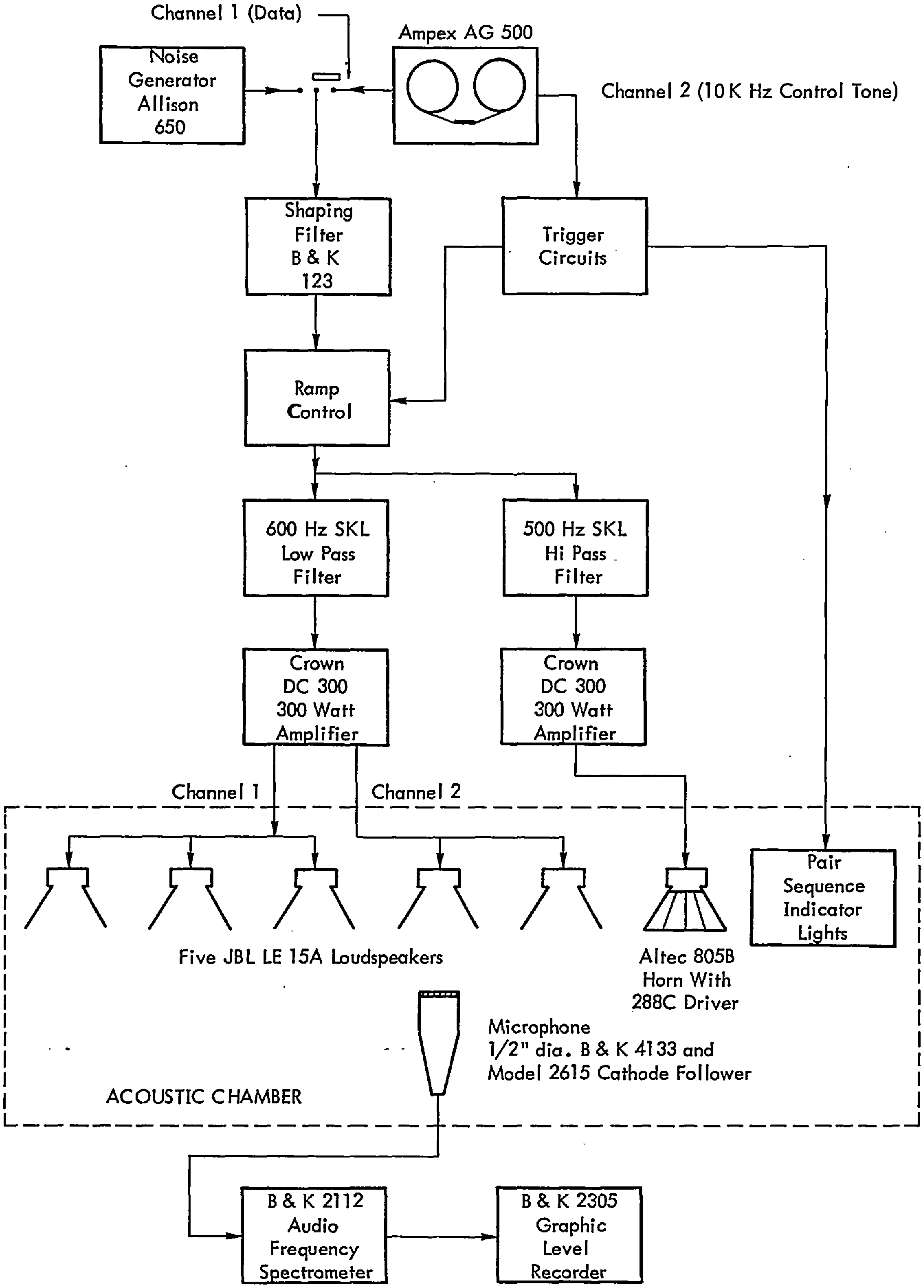


Figure 29. Sound Replay Instrumentation.

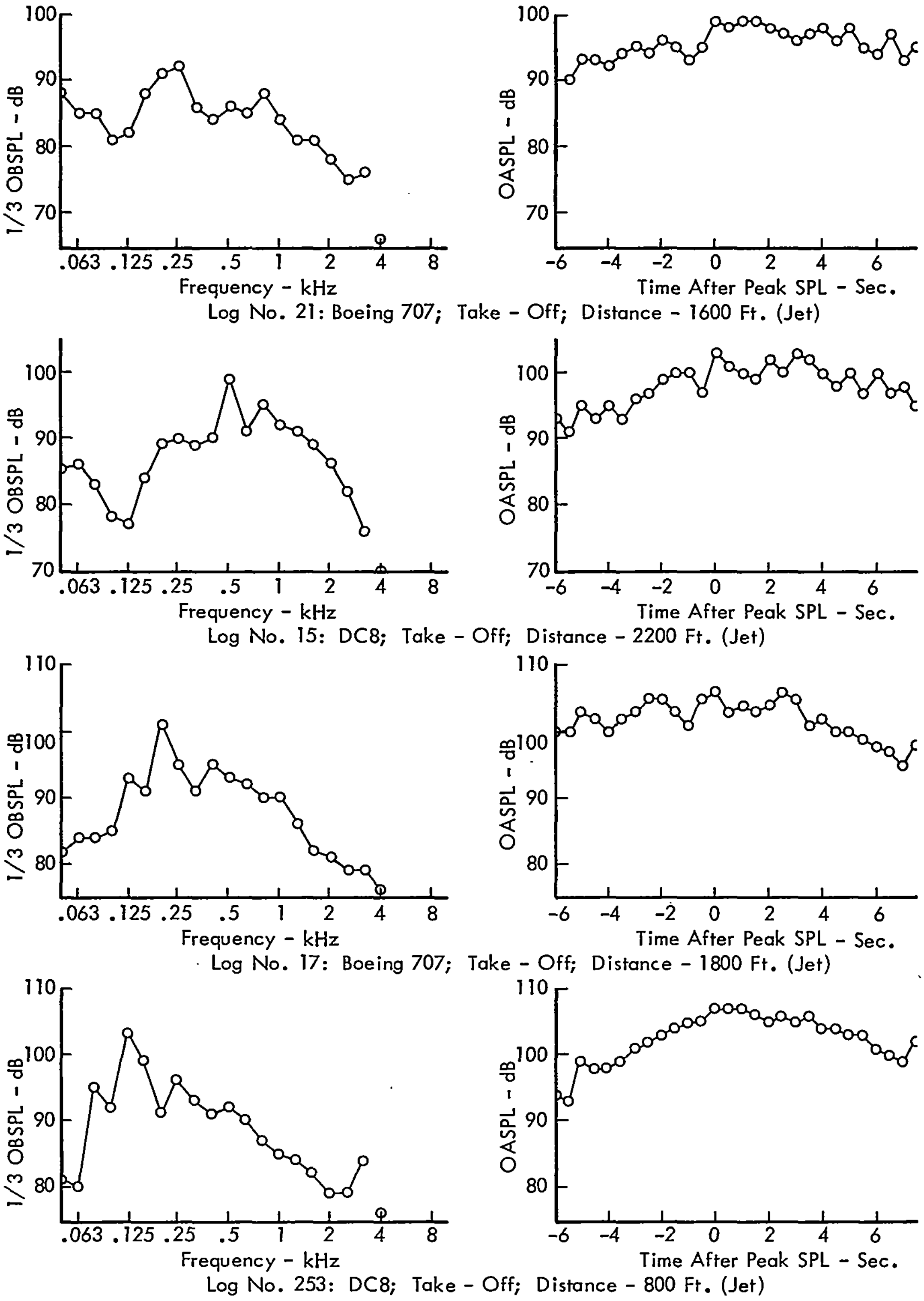
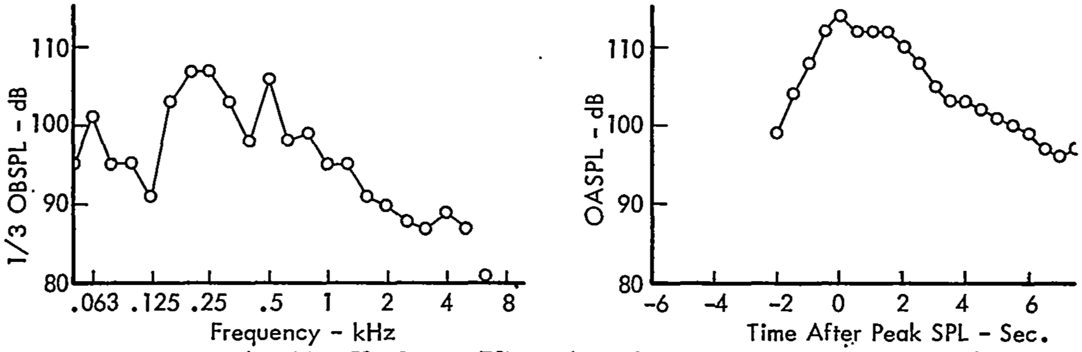
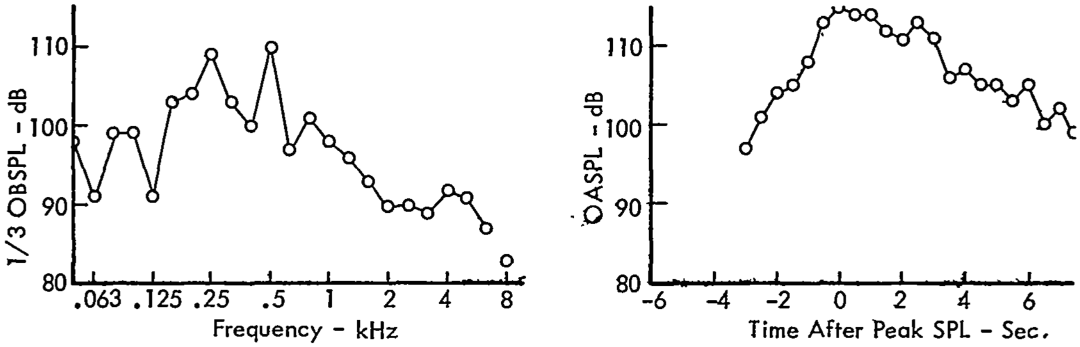


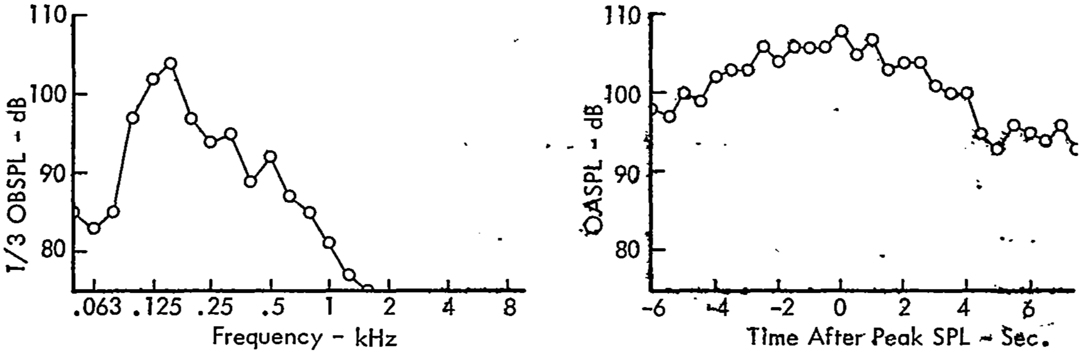
Figure 30. Time Histories and Spectra of Flyover Sounds Studied.



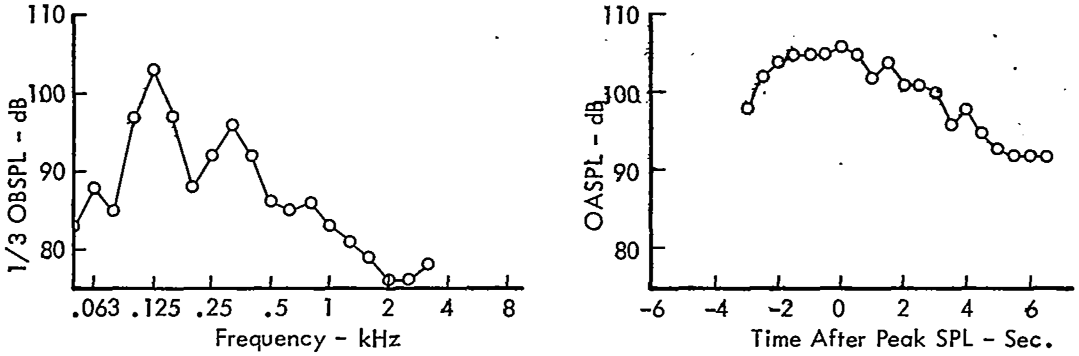
Log No. 53: Boeing 727; Take - Off; Estd. Distance - 200 Ft. (Jet)



Log No. 55: Boeing 727; Take - Off; Estd. Distance - 200 Ft. (Jet)

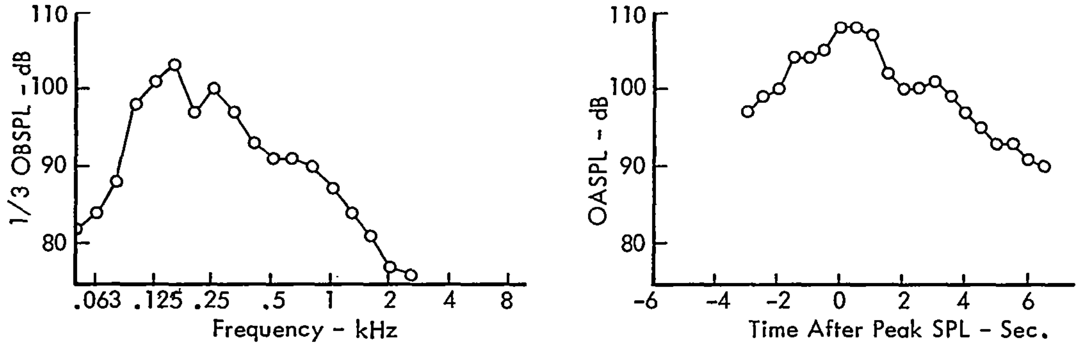


Log No. 13: Boeing 727; Take - Off; Distance - 1200 Ft. (Jet)

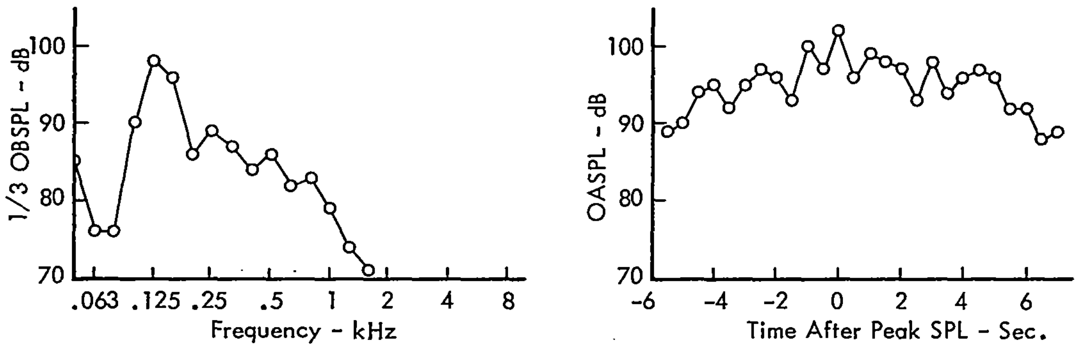


Log No. 255: DC8; Take - Off; Distance - 580 Ft. (Jet)

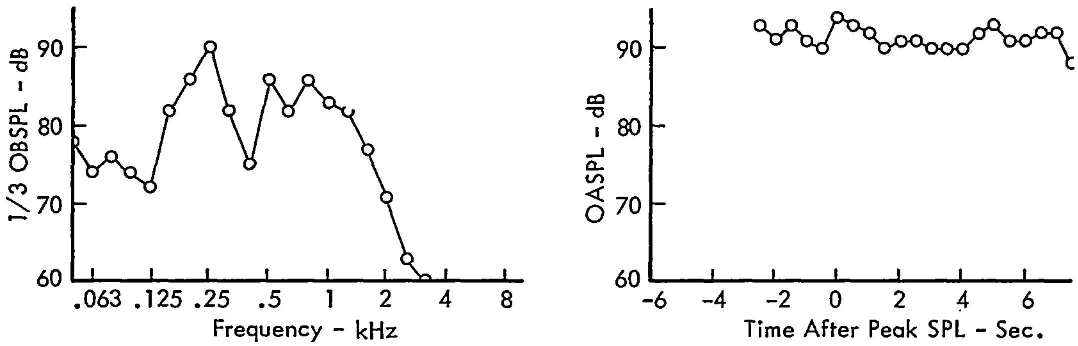
Figure 30. (Continued).



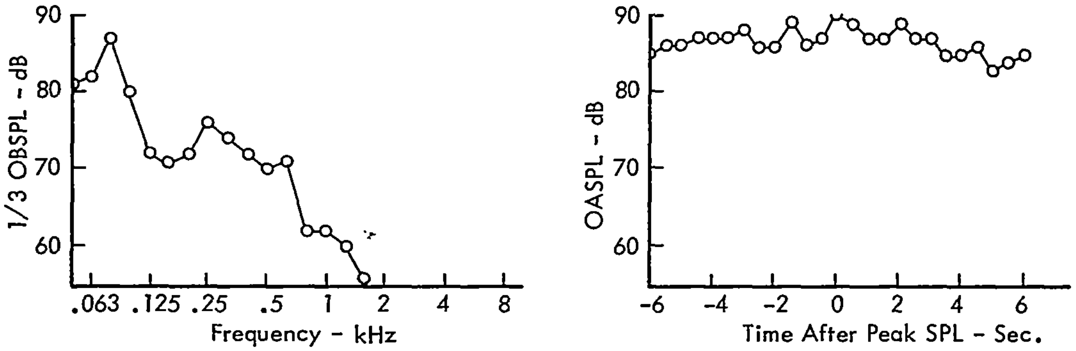
Log No. 256: Boeing 727; Take - Off; Distance - 630 Ft. (Jet)



Log No. 257: Boeing 727; Take - Off; Distance - 1300 Ft. (Jet)

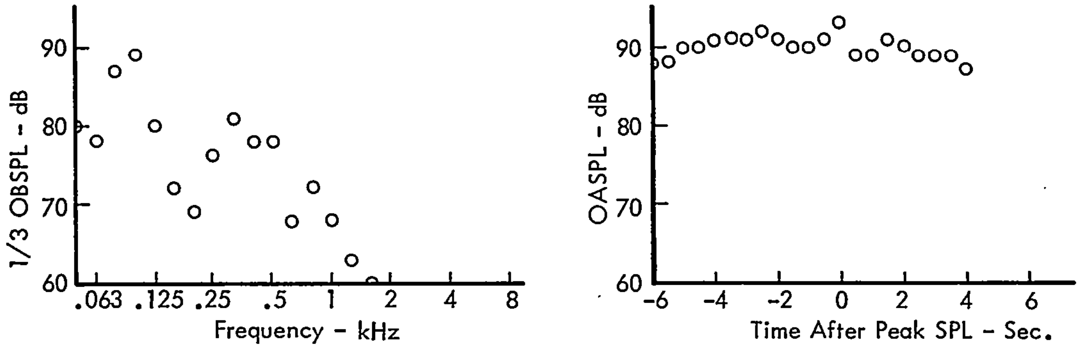


Log No. 261: Boeing 727; Take - Off; Distance - 1850 (Jet)

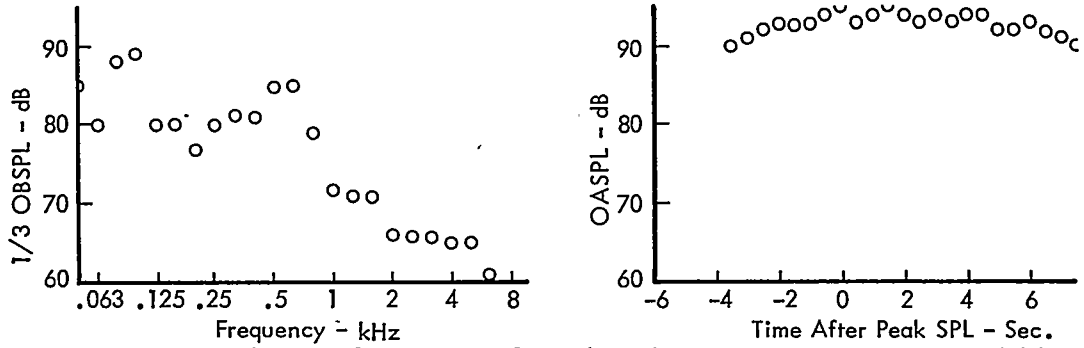


Log No. 266: Boeing 727; Take - Off; Distance - 1750 Ft. (Jet)

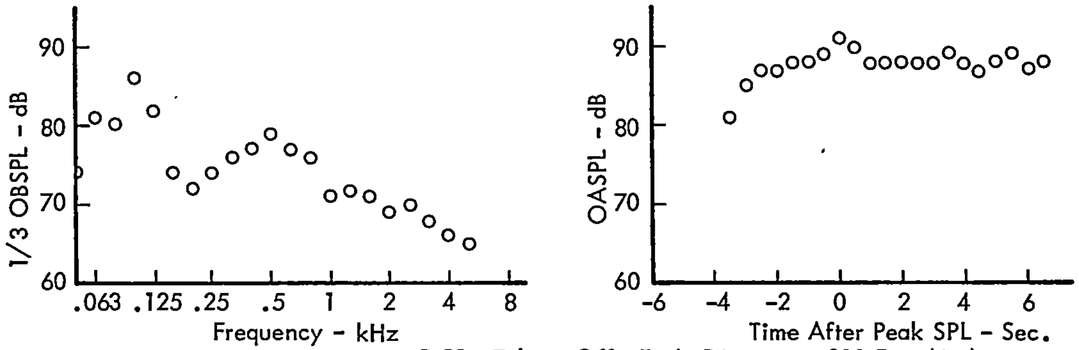
Figure 30. (Continued).



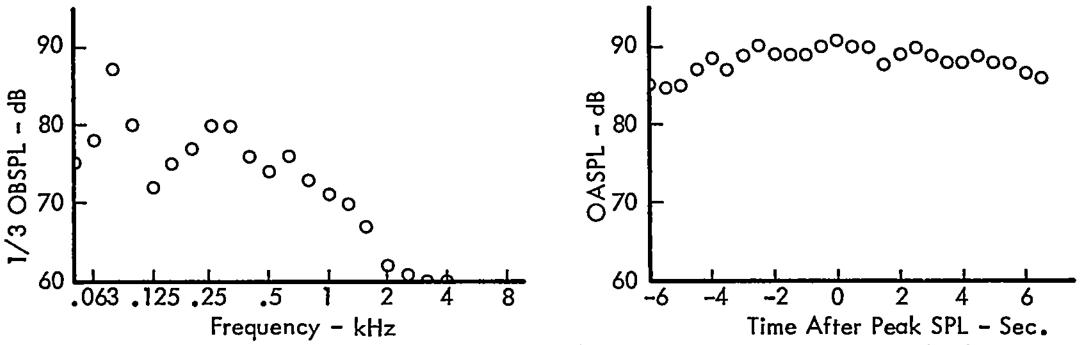
Log No. 284: DC9 ; Take - Off; Estd. Distance - 200 Ft. (Jet)



Log No. 285: Boeing 727; Take - Off; Estd. Distance - 200 Ft. (Jet)

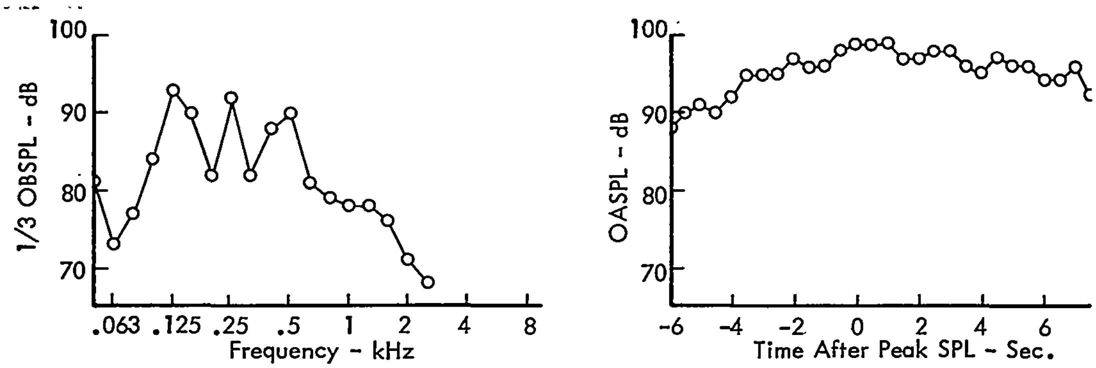


Log No. 286: DC9; Take - Off; Estd. Distance - 200 Ft. (Jet)

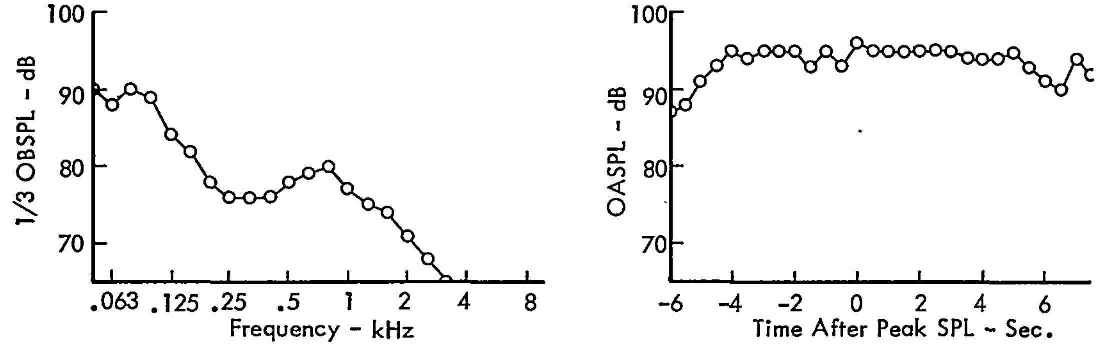


Log No. 287: Boeing 727; Take - Off; Distance - 200 Ft. (Jet)

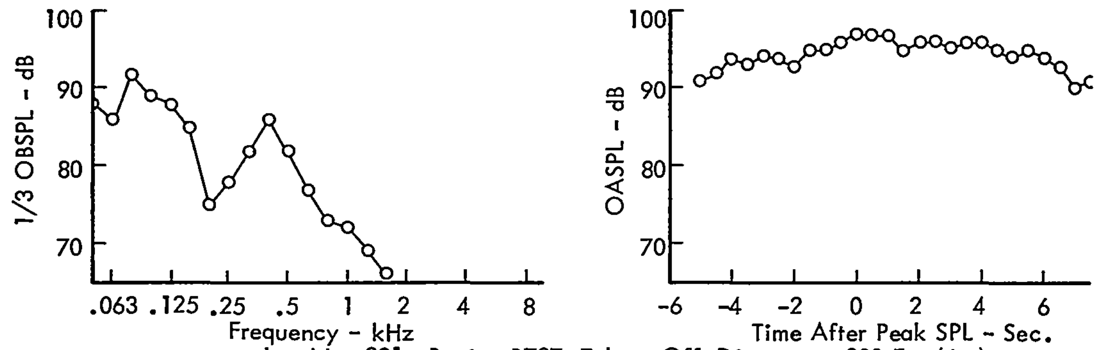
Figure 30. (Continued).



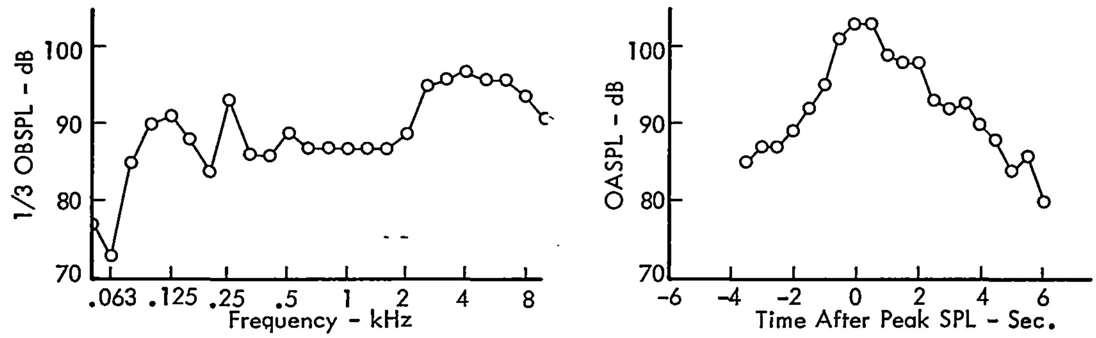
Log No. 9: DC9; Take - Off; Distance 1170 Ft. (Jet)



Log No. 290: Boeing B727; Take - Off; Distance ~200 Ft. (Jet)

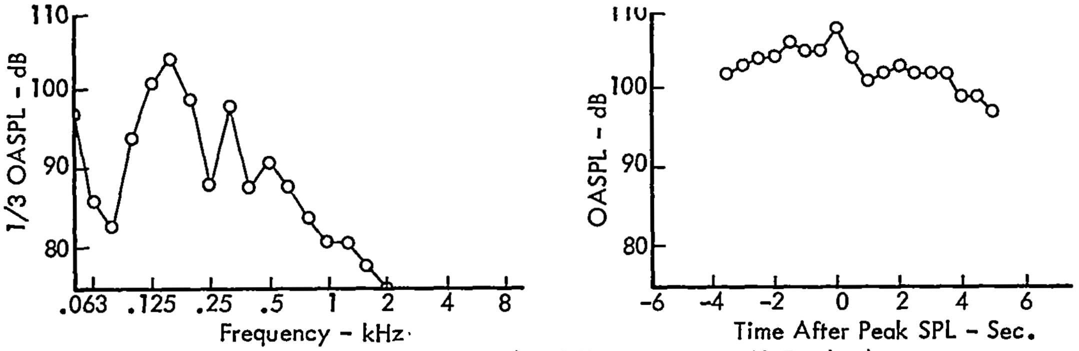


Log No. 291: Boeing B727; Take - Off; Distance ~ 200 Ft. (Jet)

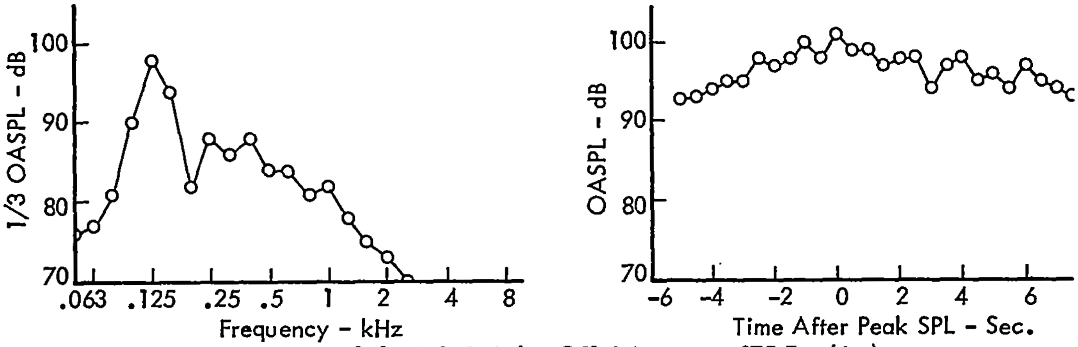


Log No. 30: DC-9; Landing; Distance - 220 Ft. (Jet)

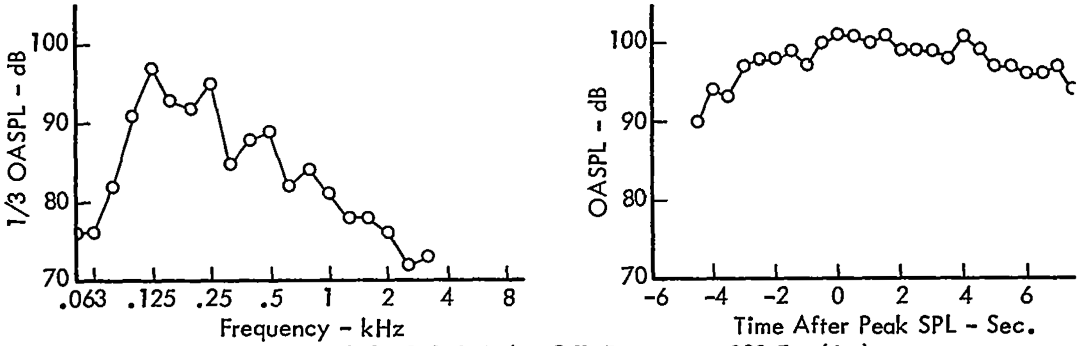
Figure 30. (Continued).



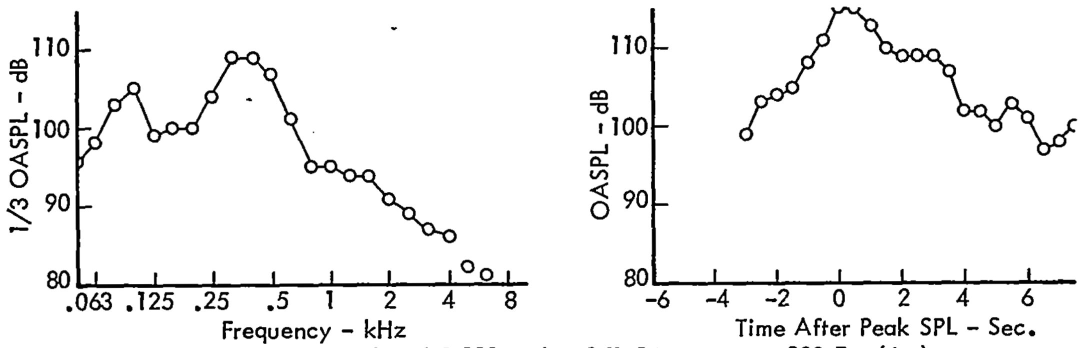
Log No. 250: DC-9; Take-Off; Distance - 660 Ft. (Jet)



Log No. 251: DC-9; Take-Off; Distance - 875 Ft. (Jet)

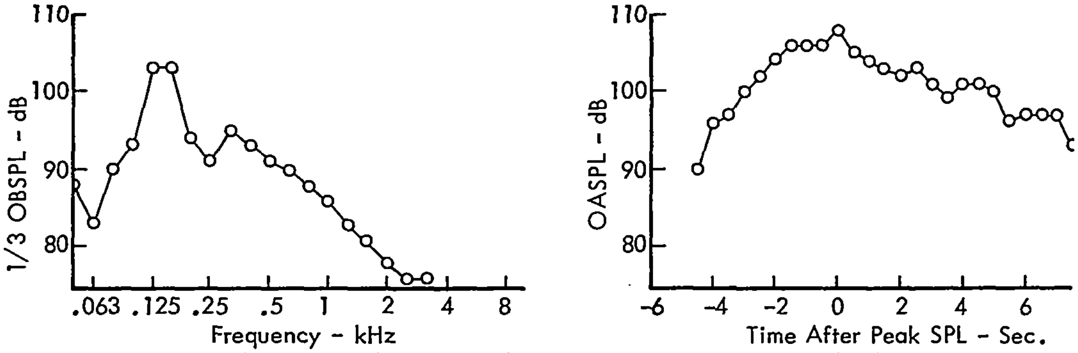


Log No. 252: DC-9; Take-Off; Distance - 930 Ft. (Jet)

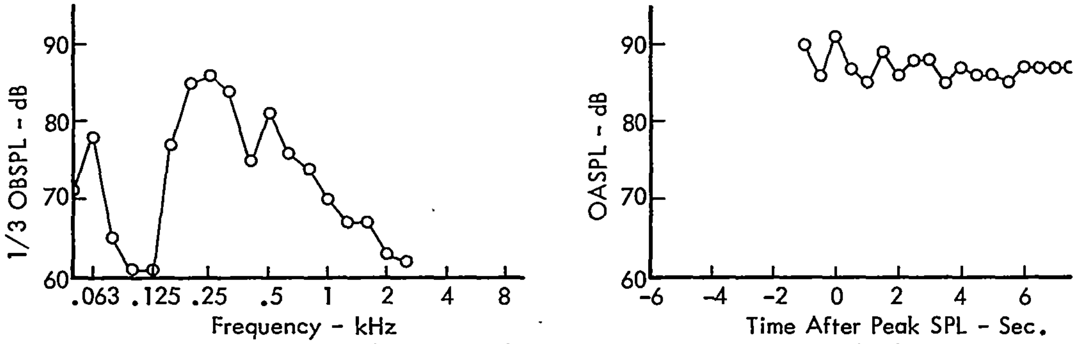


Log No. 54: BAC 111; Take-Off; Distance - ~ 200 Ft. (Jet)

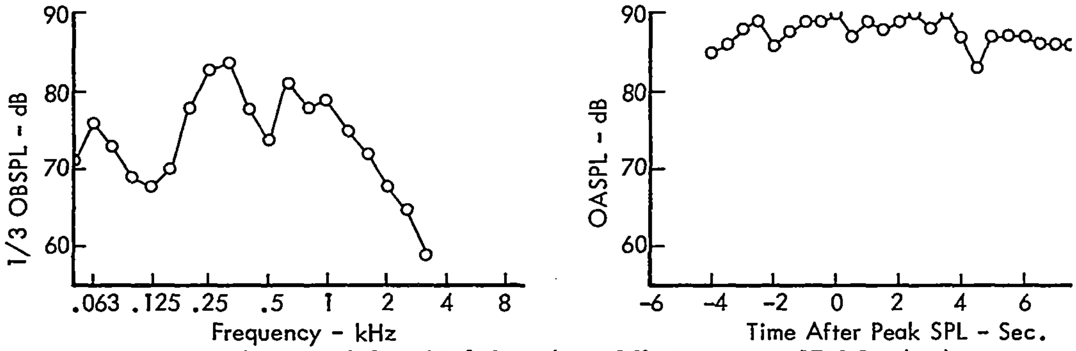
Figure 30. (Continued).



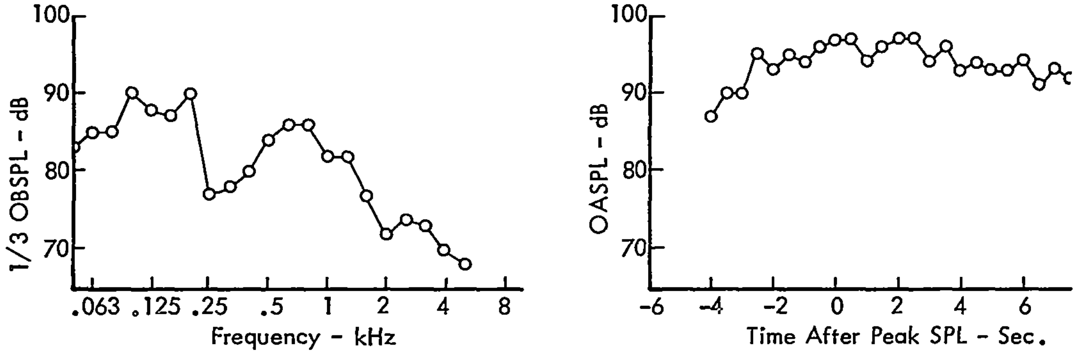
Log No. 254: DC9; Take - Off; Distance - 570 Ft. (Jet)



Log No. 262: DC9; Take - Off; Distance - 1800 Ft. (Jet)

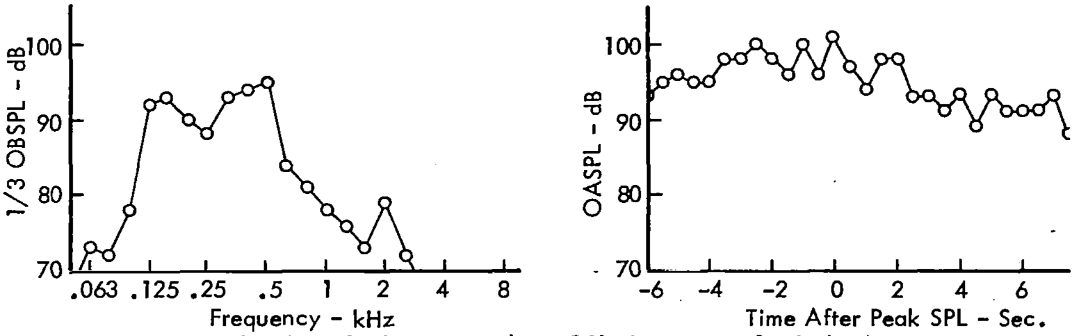


Log No. 263: BAC 111; Take - Off; Distance - 1750 Ft. (Jet)

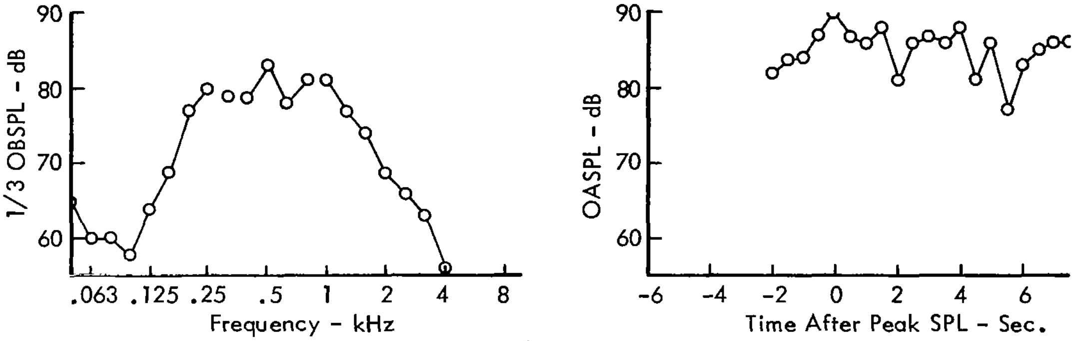


Log No. 288: BAC 111; Take - Off; Estd. Distance - 200 Ft. (Jet)

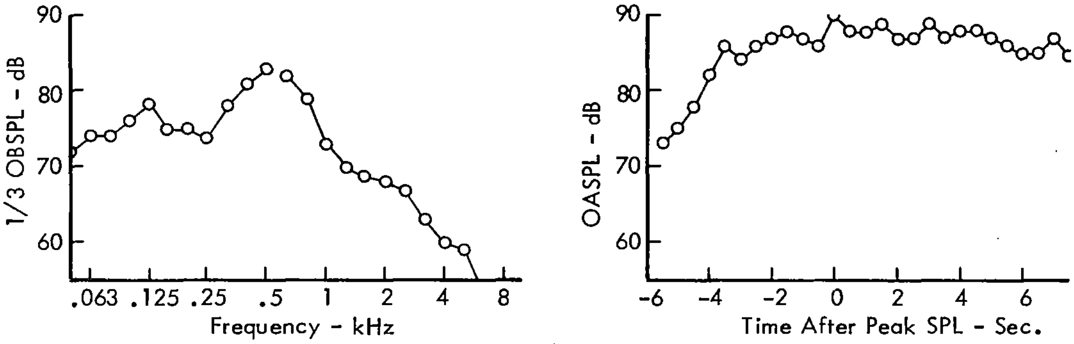
Figure 30. (Continued).



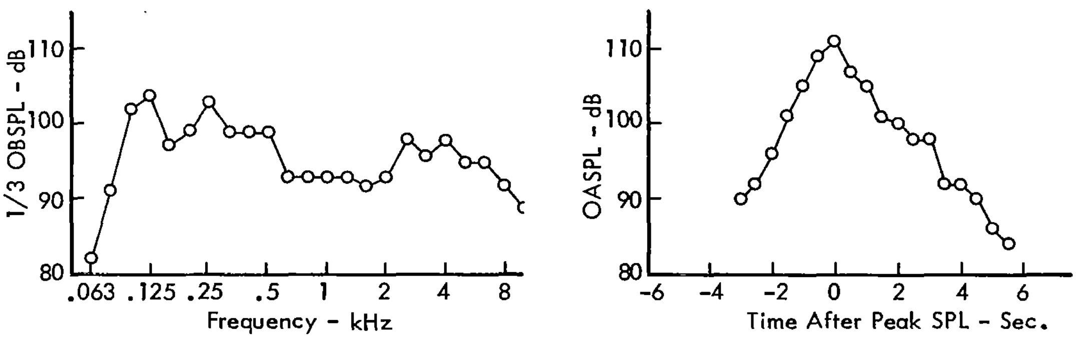
Log No. 8: Learjet; Take - Off; Distance - 1560 (Jet)



Log No. 264: HS 125; Take - Off; Distance - 1800 Ft. (Jet)

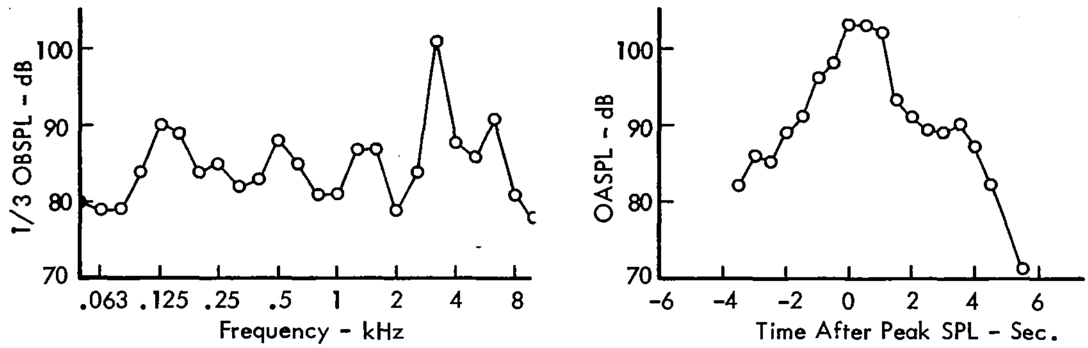


Log No. 289: BAC 111; Take - Off; Estd. Distance - 200 Ft. (Jet)

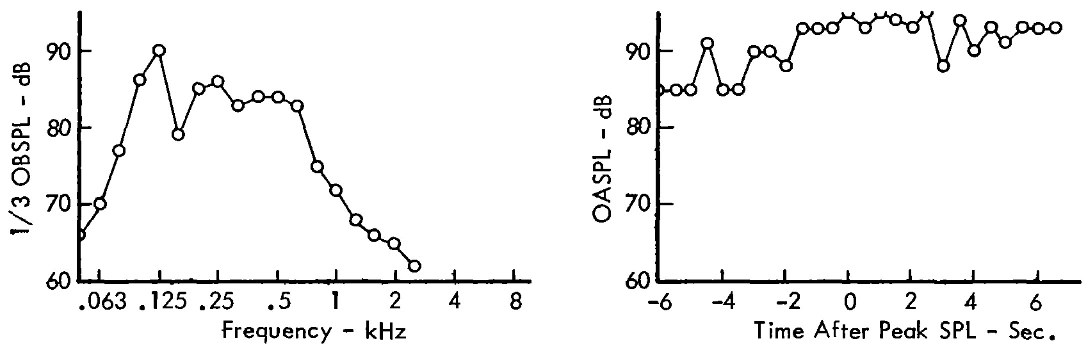


Log No. 31: Boeing 727; Landing; Distance - 210 Ft. (Jet)

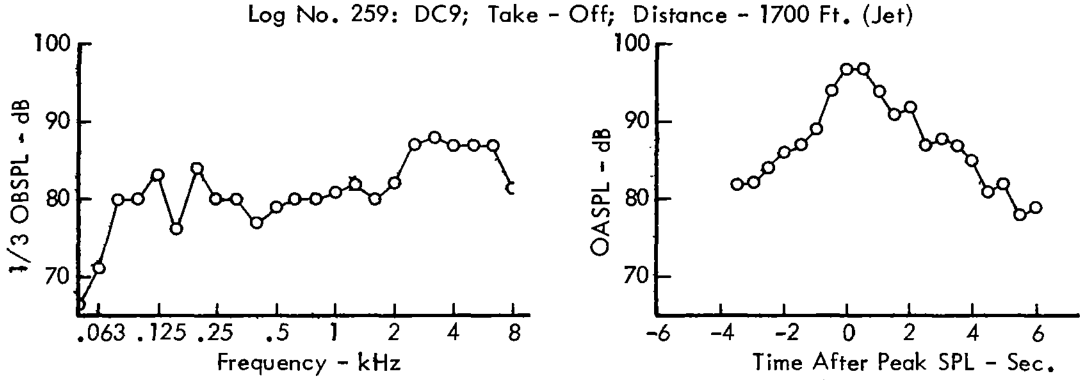
Figure 30. (Continued).



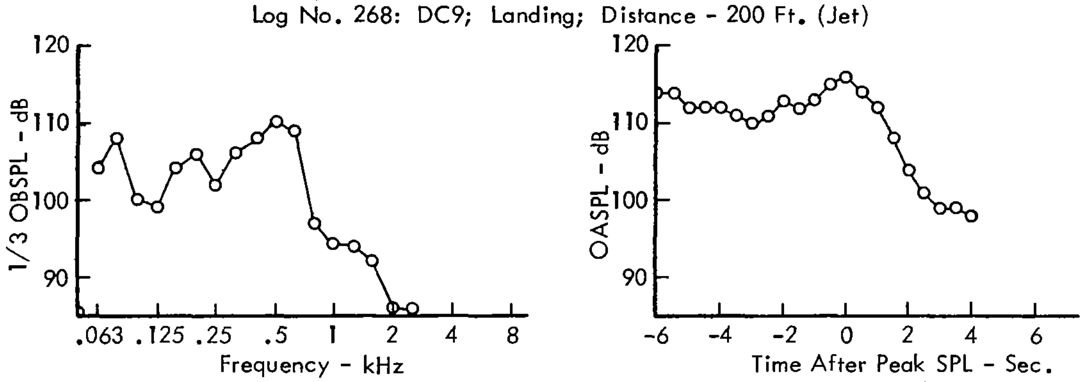
Log No. 88: Boeing 707; Landing; Distance - 200 Ft. (Jet)



Log No. 259: DC9; Take - Off; Distance - 1700 Ft. (Jet)

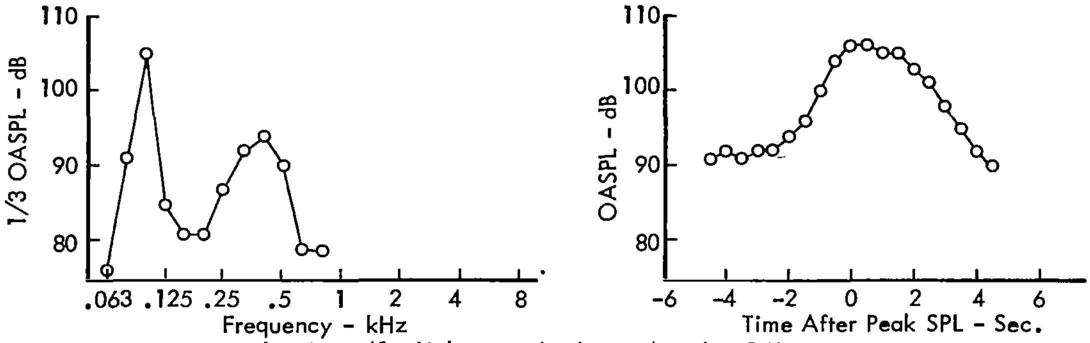


Log No. 268: DC9; Landing; Distance - 200 Ft. (Jet)

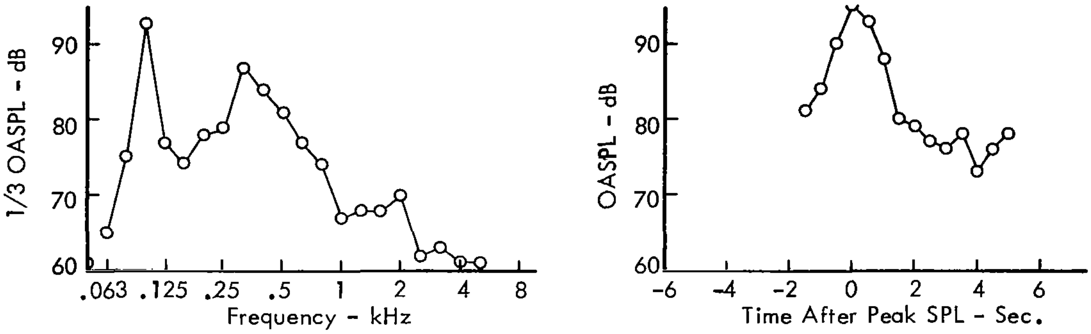


Log No. 65: XC - 142; Landing; Estd. Distance - 500 Ft. (Turboprop)

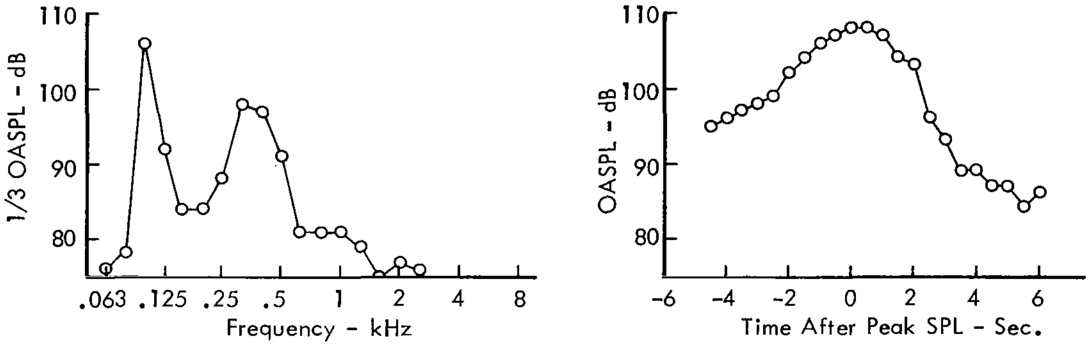
Figure 30. (Continued).



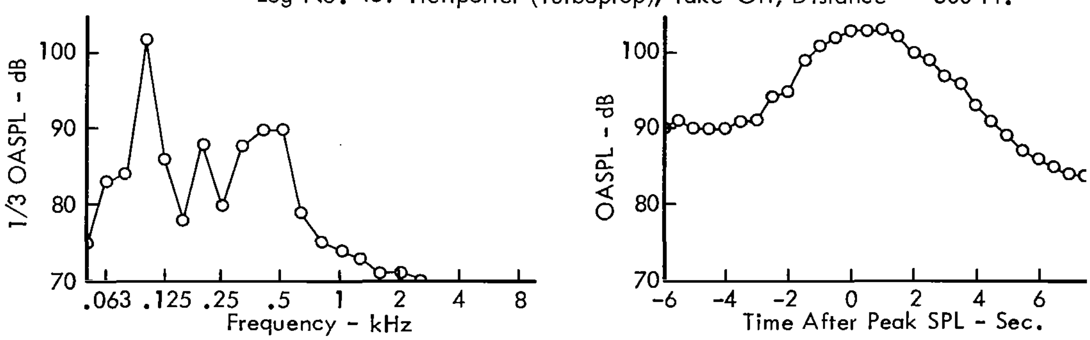
Log No. 41: Heliporter (Turboprop); Take-Off; Distance - ~ 300 Ft.



Log No. 42: Heliporter (Turboprop); Flyover; Distance - ~ 400 Ft.

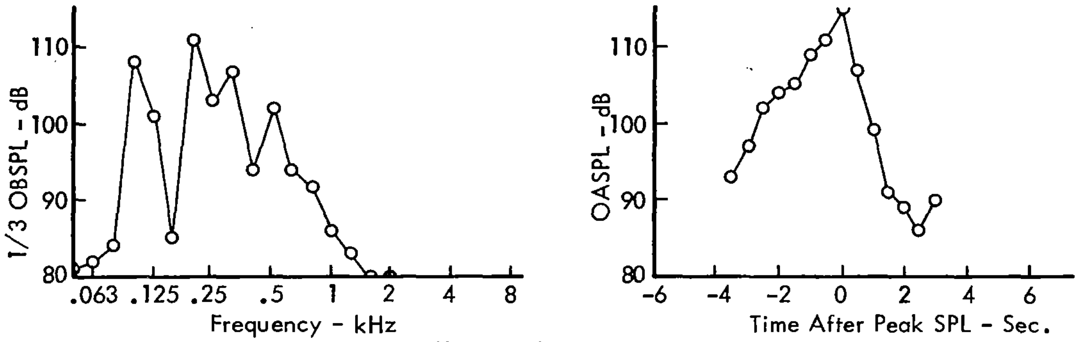


Log No. 43: Heliporter (Turboprop); Take-Off; Distance - ~ 300 Ft.

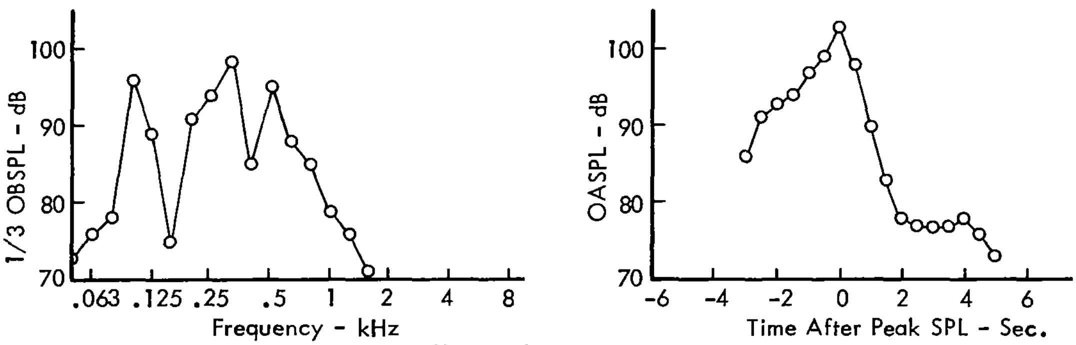


Log No. 44: Heliporter (Turboprop); Take-Off; Distance - ~ 300 Ft.

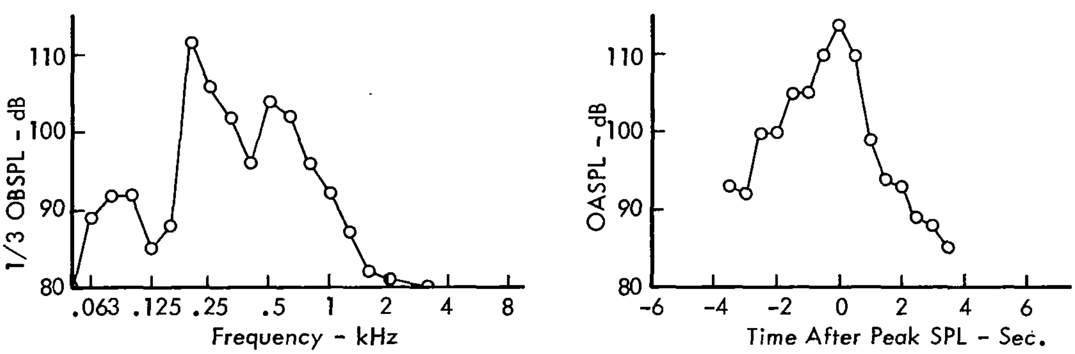
Figure 30. (Continued).



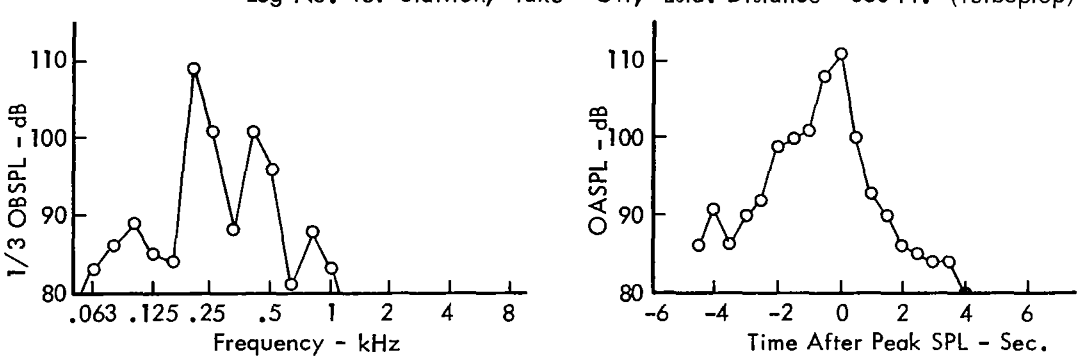
Log No. 46: Stallion; Take - Off; Estd. Distance - 350 Ft. (Turboprop)



Log No. 47: Stallion; Flyover; Estd. Distance - 350 Ft. (Turboprop)

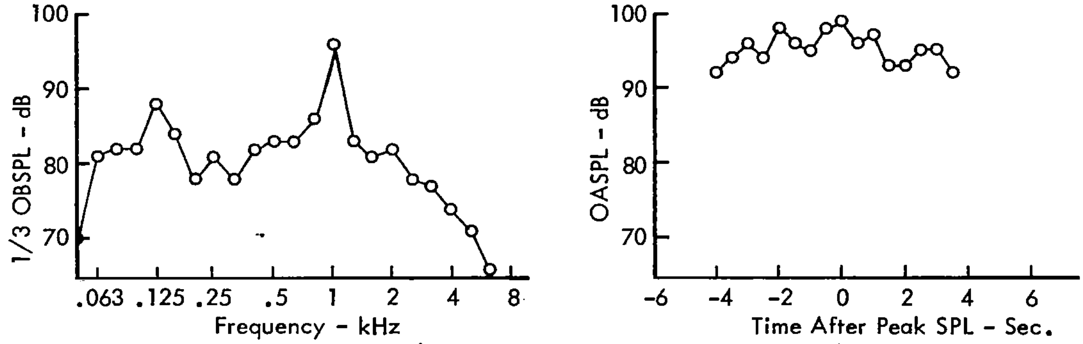


Log No. 48: Stallion; Take - Off, Estd. Distance - 350 Ft. (Turboprop)

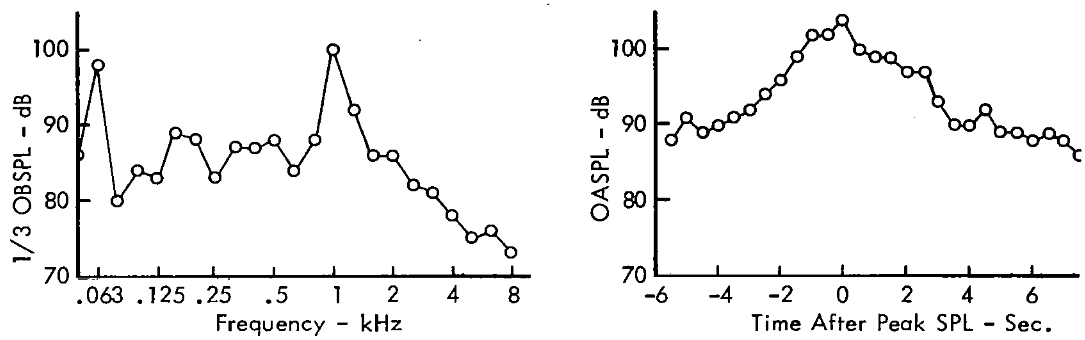


Log No. 49: Stallion; Flyover; Estd. Distance - 350 Ft. (Turboprop)

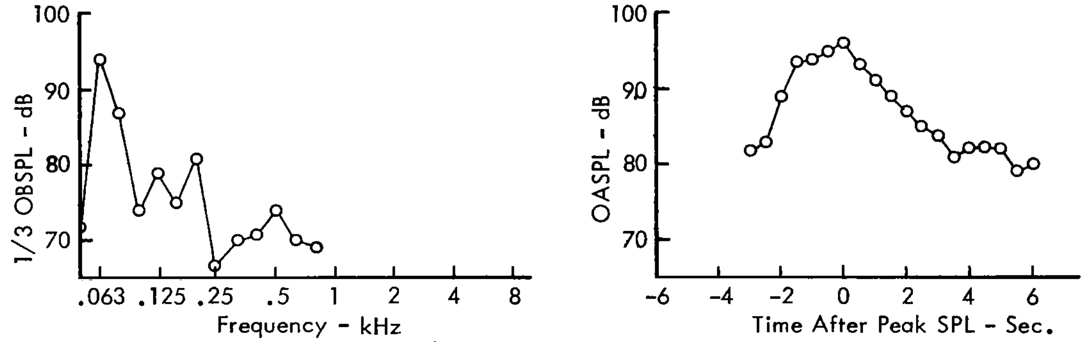
Figure 30. (Continued).



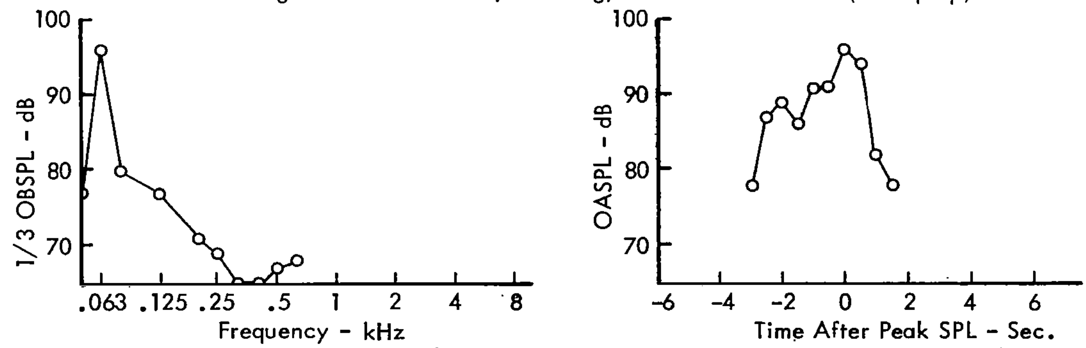
Log No. 78: Electra; **Landing; Distance - 620 Ft.** (Turboprop)



Log No. 81: Electra; Landing; Distance - 310 Ft. (Turboprop)

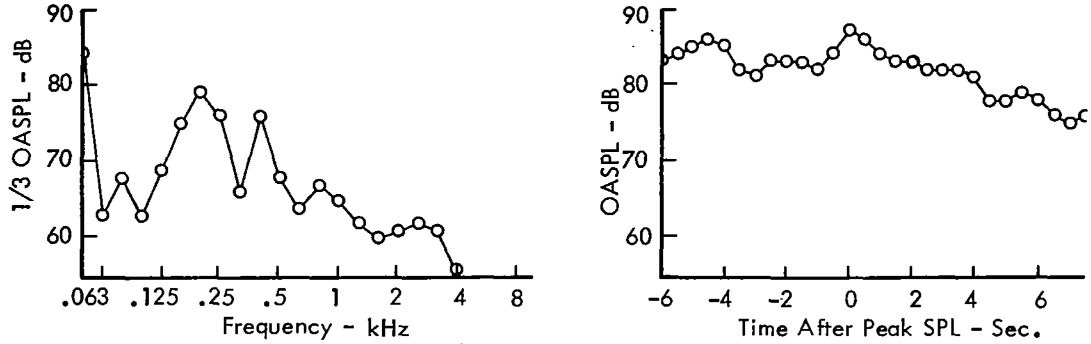


Log No. 280: Electra; Landing; Distance - 400 Ft. (Turboprop)

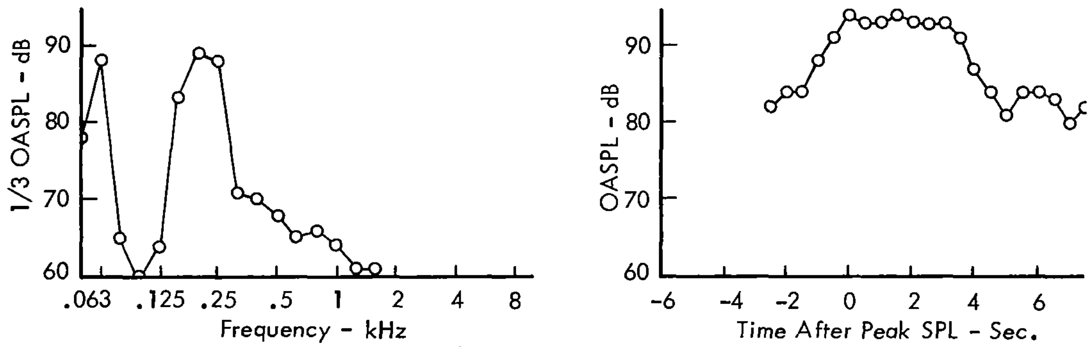


Log No. 281: Electra; Take - Off; Distance - 600 Ft. (Turboprop)

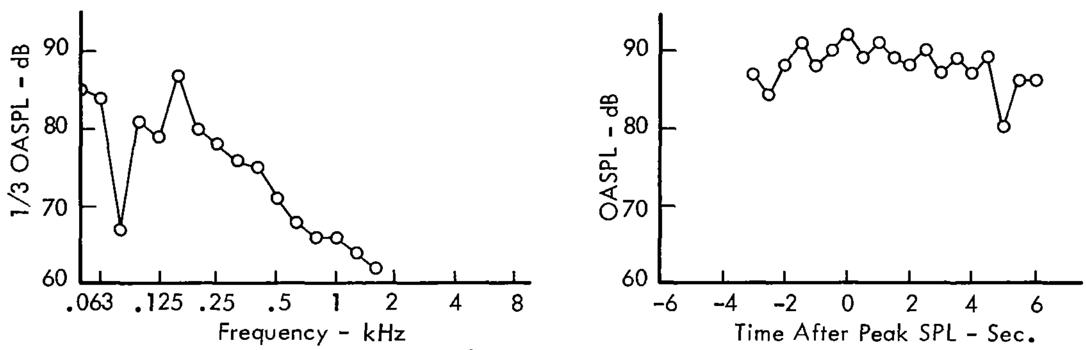
Figure 30. (Continued).



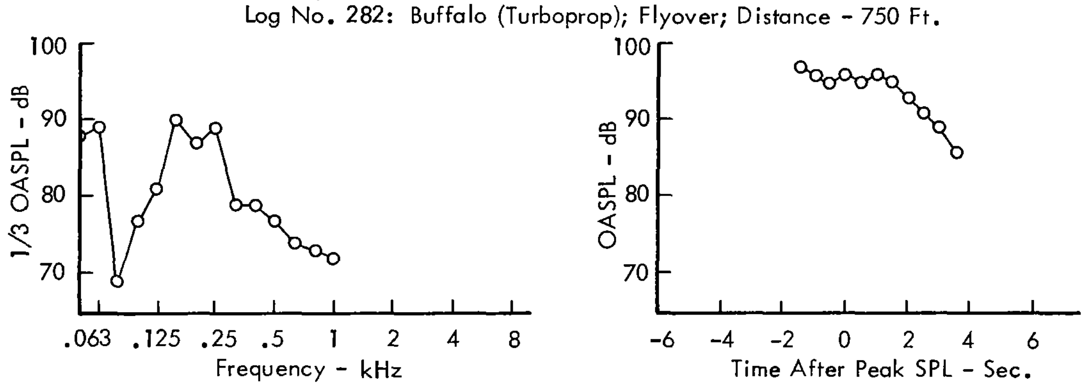
Log No. 89: Buffalo (Turboprop); Flyover; Distance - 750 Ft.



Log No. 90: Buffalo (Turboprop); Flyover; Distance - 750 Ft.

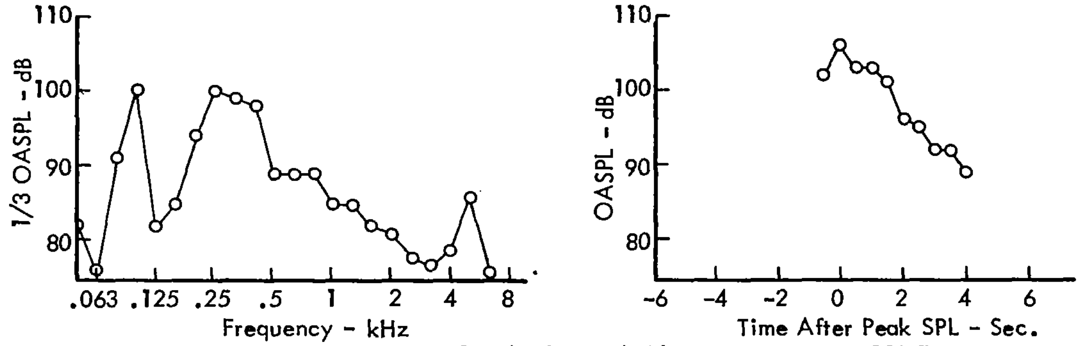


Log No. 282: Buffalo (Turboprop); Flyover; Distance - 750 Ft.

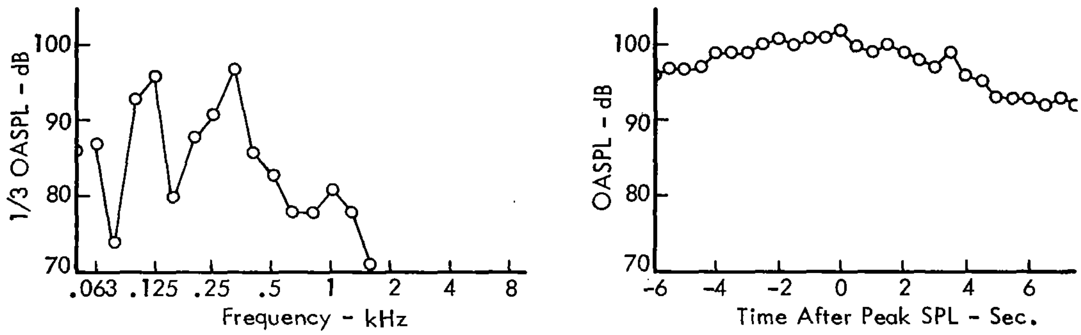


Log No. 283: Buffalo (Turboprop); Flyover; Distance - 750 Ft.

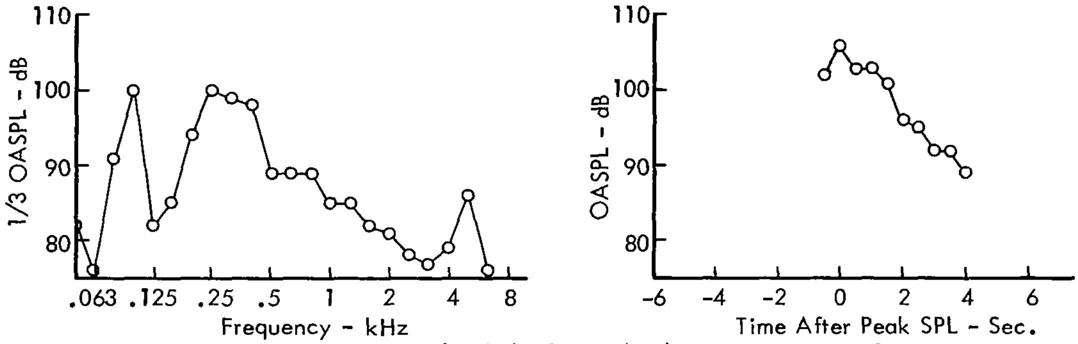
Figure 30. (Continued).



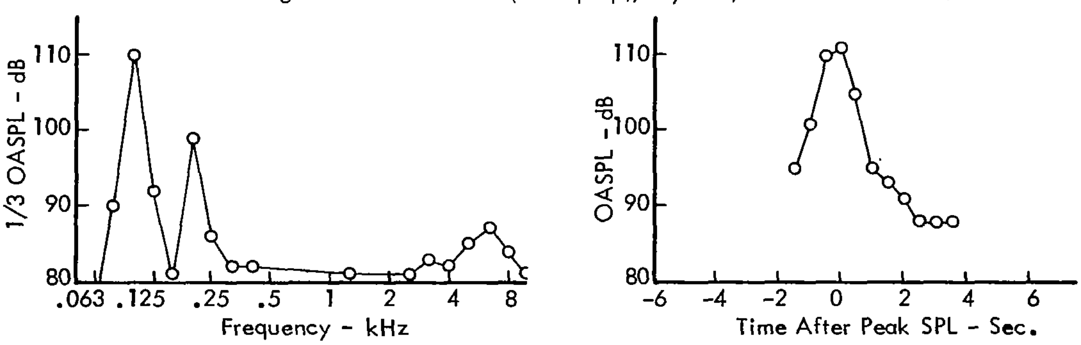
Log No. 61: MD 188 (Turboprop); Flyover; Distance - 320 Ft.



Log No. 63: MD 188 (Turboprop); Take-Off; Distance - 500 Ft.

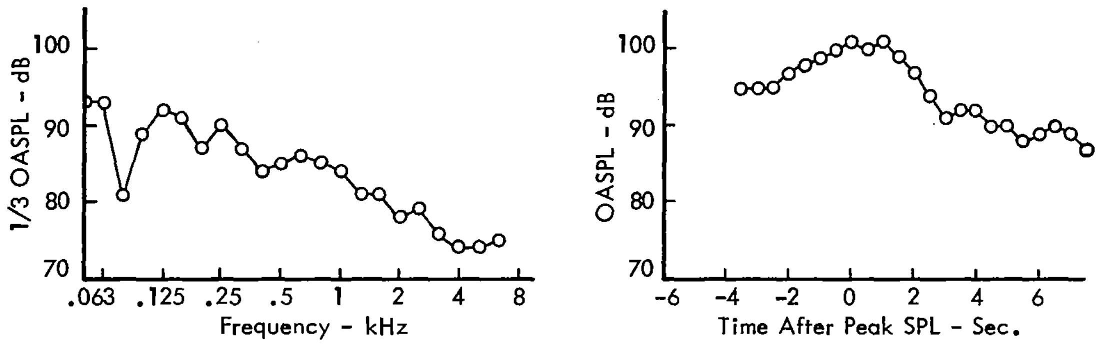


Log No. 72: Mohawk (Turboprop); Flyover; Distance - 300 Ft.

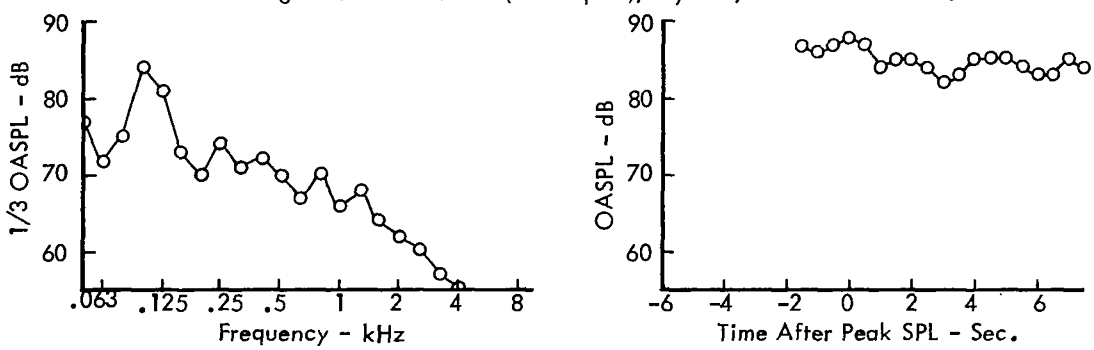


Log No. 87: Skyvan (Turboprop); Landing; Distance - ~ 500 Ft.

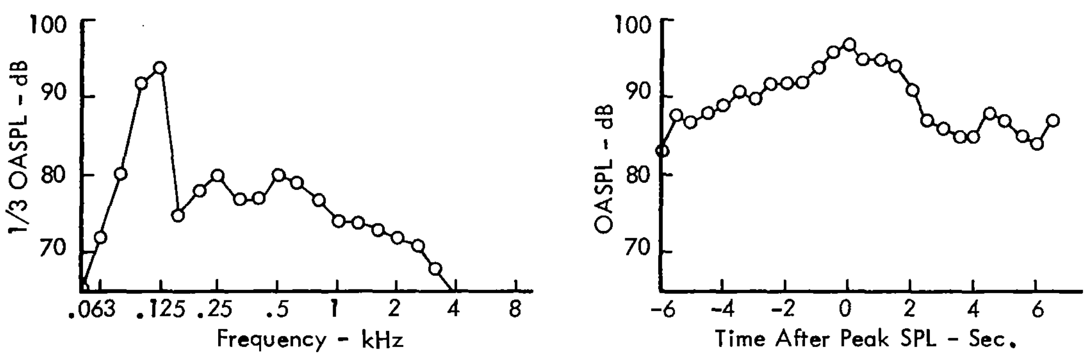
Figure 30. (Continued).



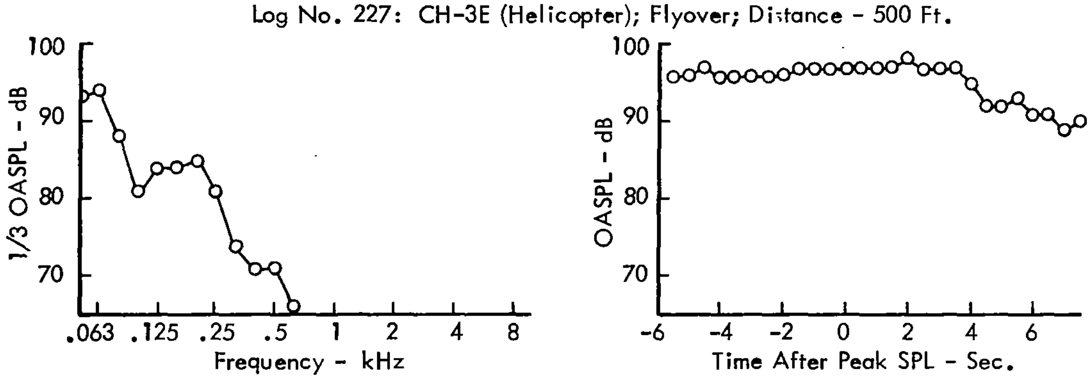
Log No. 225: CH-3E (Helicopter); Flyover; Distance - 125 Ft.



Log No. 226: CH-3E (Helicopter); Flyover; Distance - 1000 Ft.

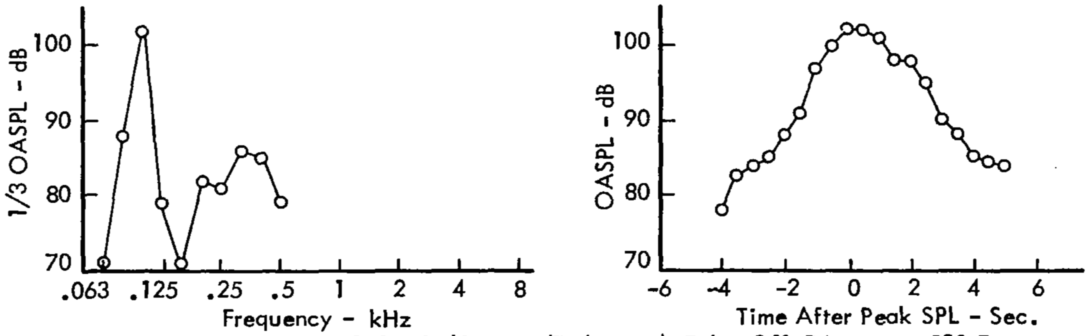


Log No. 227: CH-3E (Helicopter); Flyover; Distance - 500 Ft.

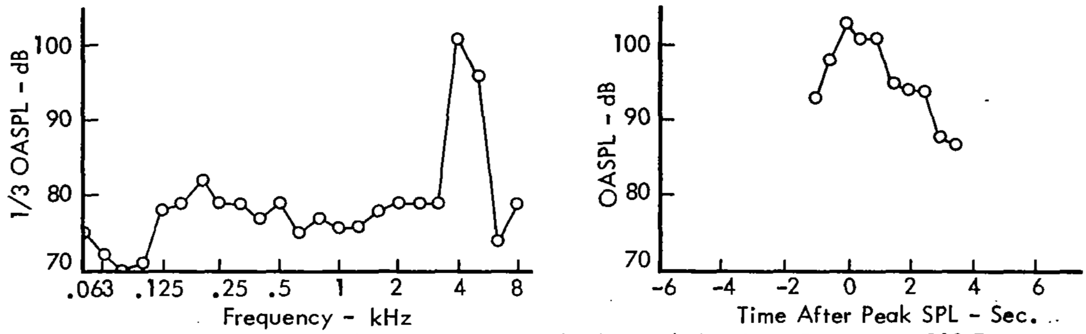


Log No. 232: UH-1F (Helicopter); Flyover; Distance - 500 Ft.

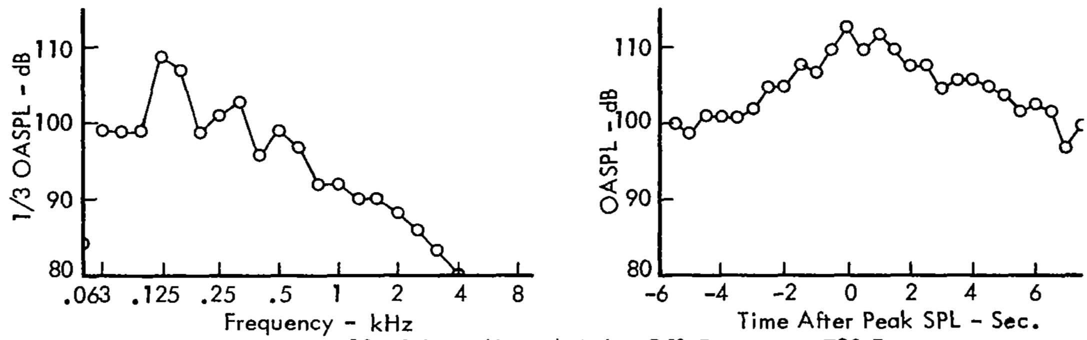
Figure 30. (Continued).



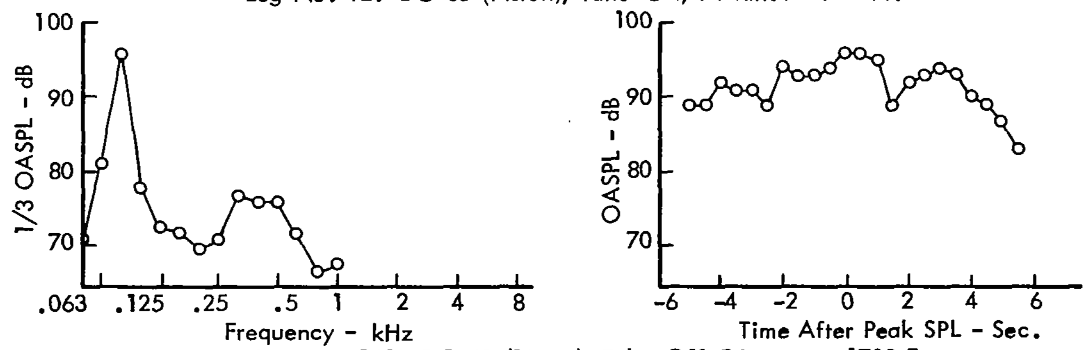
Log No. 258: Gulfstream (Turboprop); Take-Off; Distance - 930 Ft.



Log No. 267: Convair 580 (Turboprop); Landing; Distance - 200 Ft.

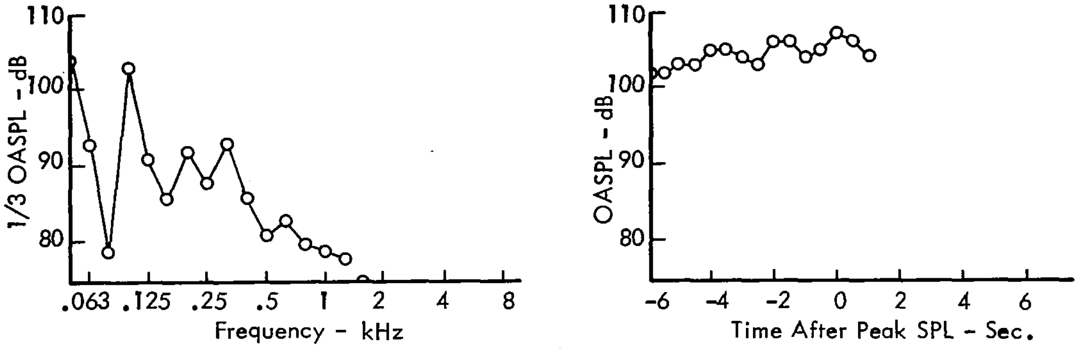


Log No. 12: DC-6B (Piston); Take-Off; Distance - 720 Ft.

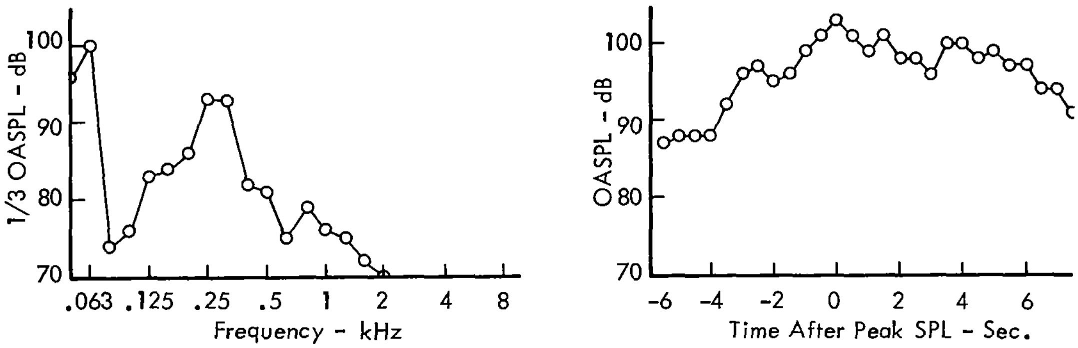


Log No. 260: DC-6B (Piston); Take-Off; Distance - 1700 Ft.

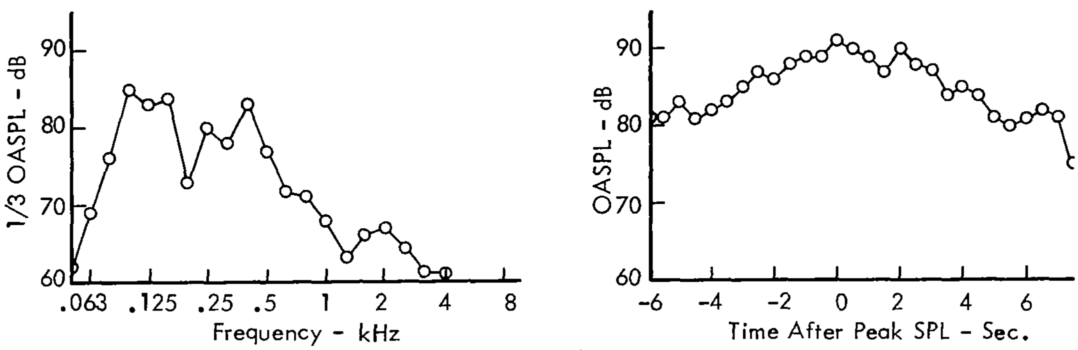
Figure 30. (Continued).



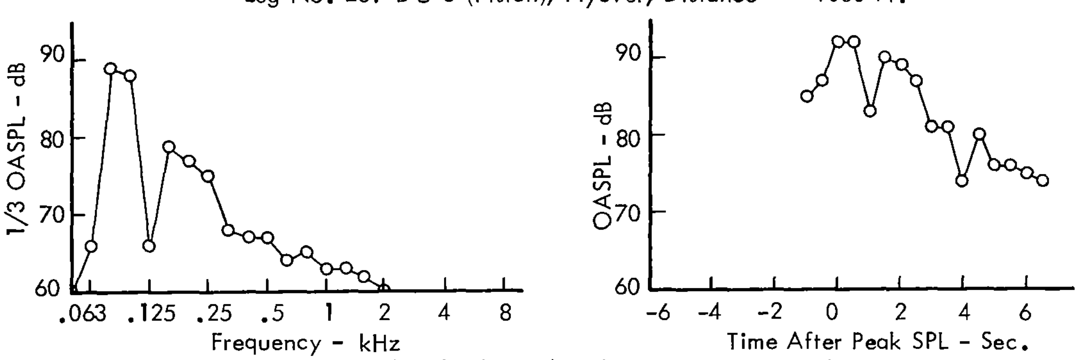
Log No. 265: C-124 (Piston); Take-Off; Distance - 1700 Ft.



Log No. 26: C-124 (Piston); Flyover; Distance - ~ 750 Ft.

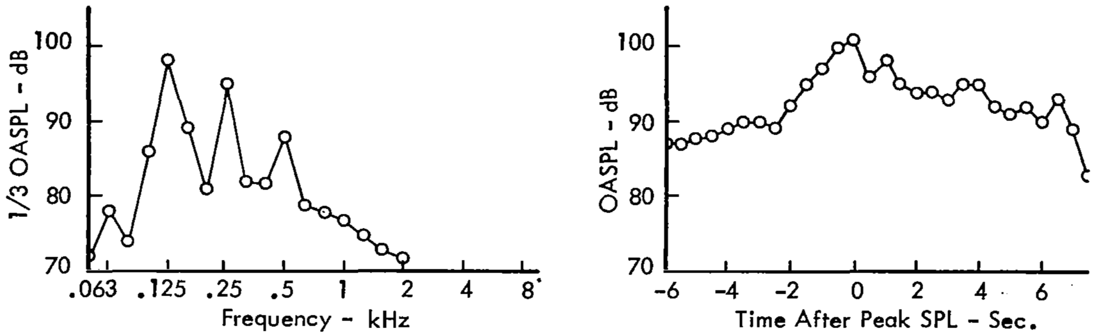


Log No. 23: DC-3 (Piston); Flyover; Distance - ~ 1000 Ft.

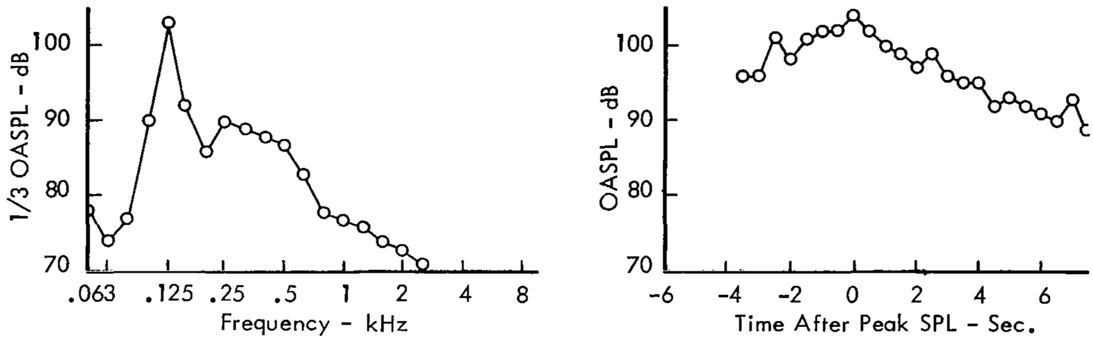


Log No. 27: Apache (Piston); Take-Off; Distance - 400 Ft.

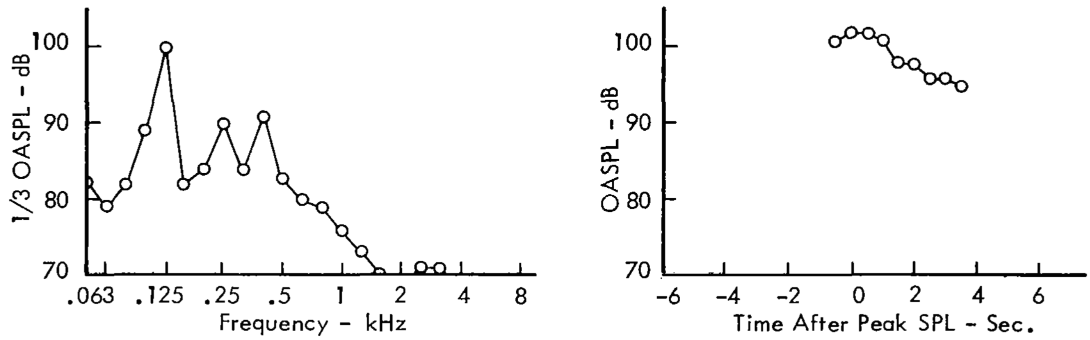
Figure 30. (Continued).



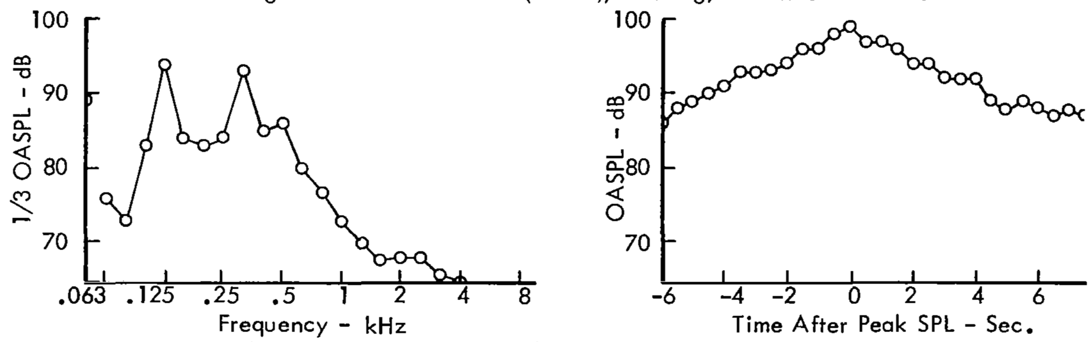
Log No. 10: Martin 404 (Piston); Take-Off; Distance - 1000 Ft.



Log No. 74: Martin 404 (Piston); Landing; Distance - 800 Ft.

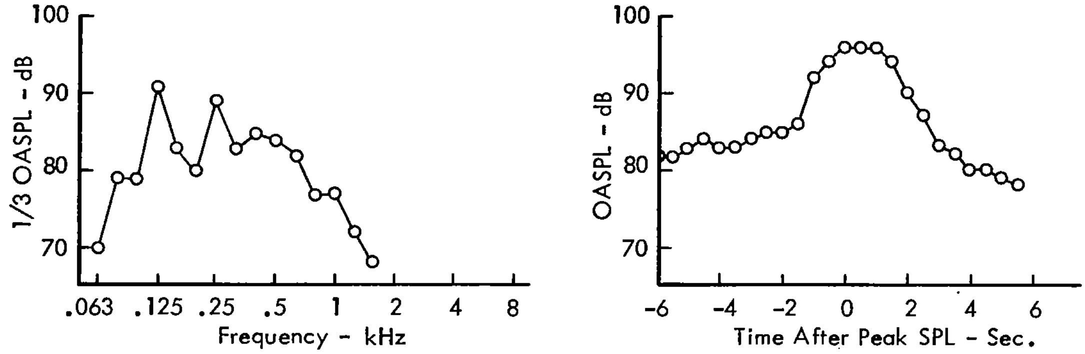


Log No. 77: Martin 404 (Piston); Landing; Distance - 650 Ft.

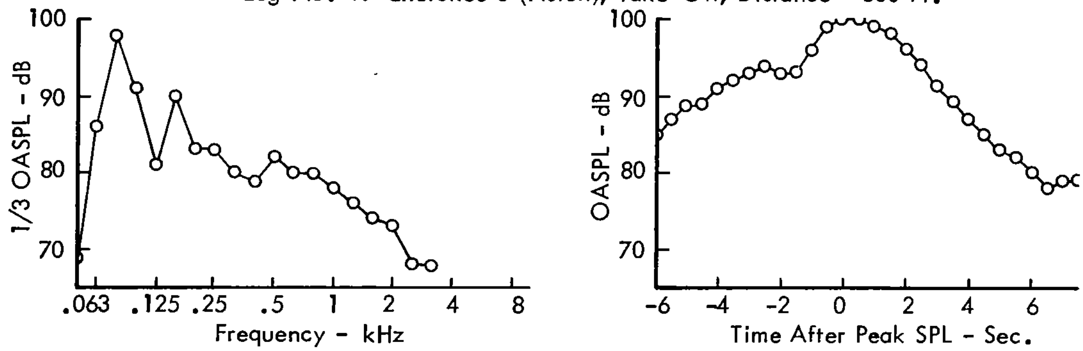


Log No. 79: Martin 404 (Piston); Landing; Distance - 700 Ft.

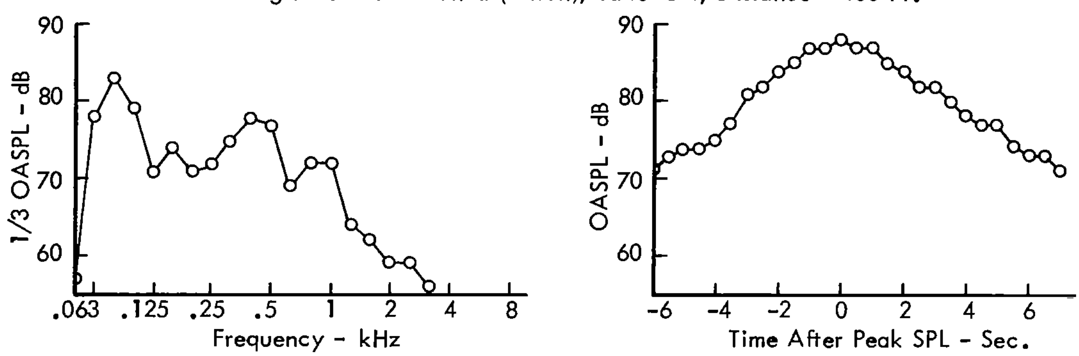
Figure 30. (Continued).



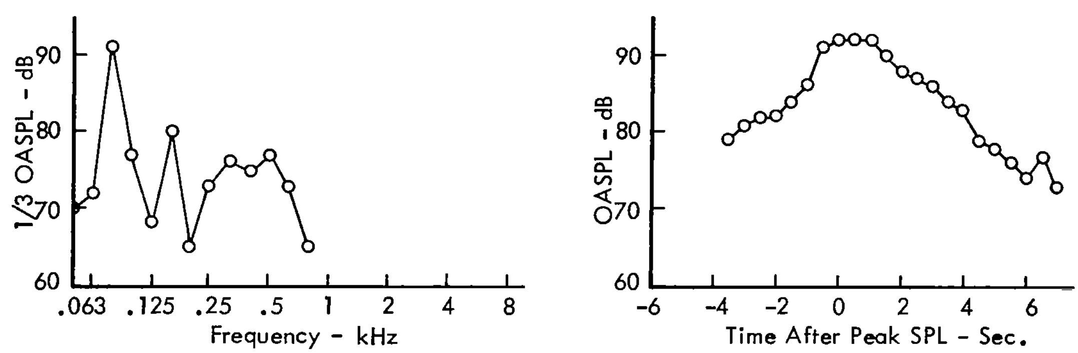
Log No. 4: Cherokee 6 (Piston); Take-Off; Distance - 300 Ft.



Log No. 20: Bonanza (Piston); Take-Off; Distance - 400 Ft.

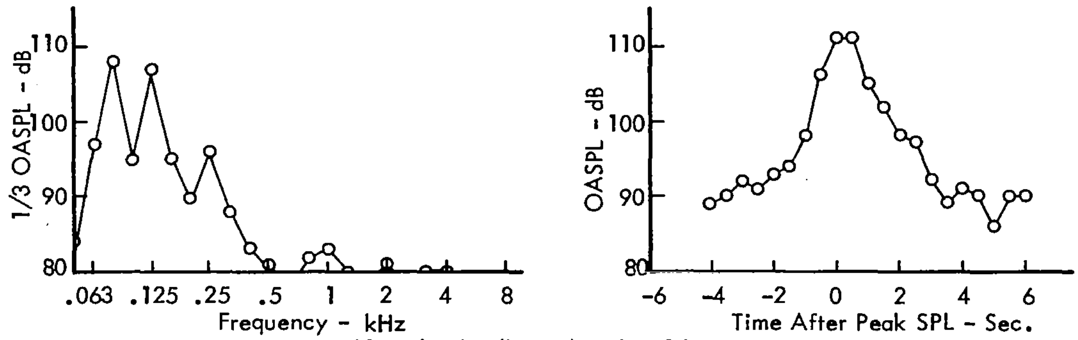


Log No. 25: Cessna 172 (Piston); Take-Off; Distance - 500 Ft.

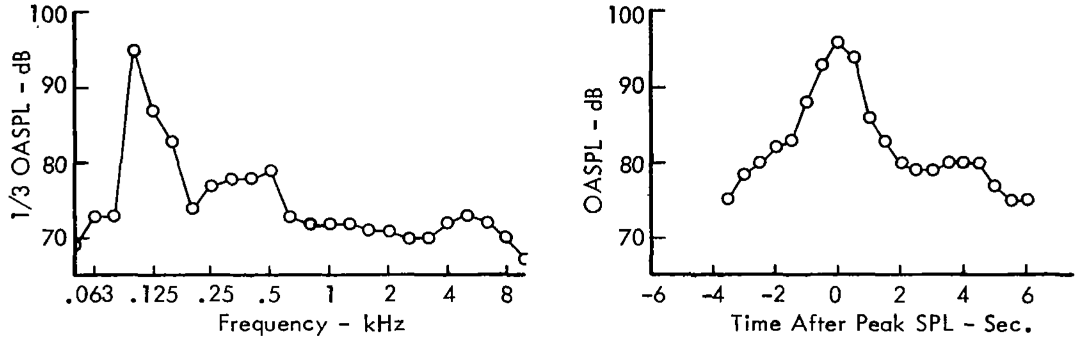


Log No. 28: Cessna 172 (Piston); Take-Off; Distance - 400 Ft.

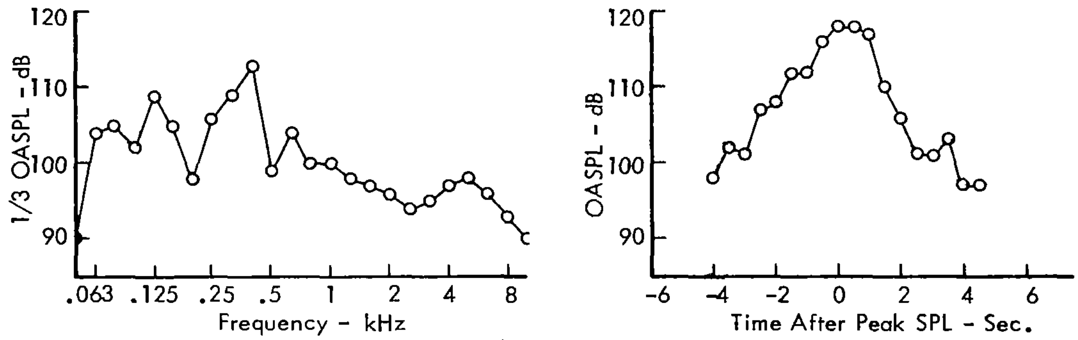
Figure 30. (Continued).



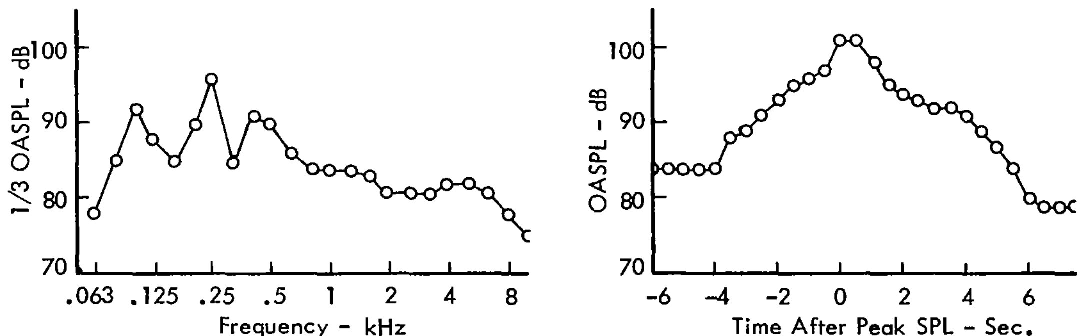
Log No. 33: Islander (Piston); Take-Off; Distance - ~ 200 Ft.



Log. No. 34: Islander (Piston); Flyover; Distance - 200 Ft.

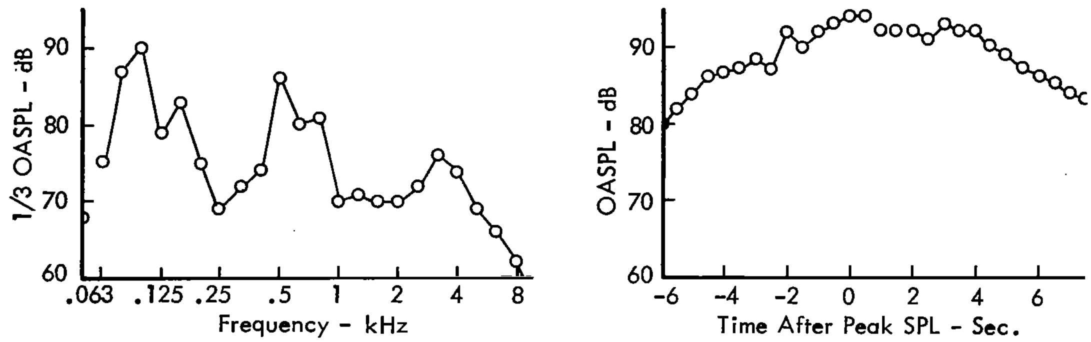


Log No. 35: Islander (Piston); Take-Off; Distance - ~ 200 Ft.

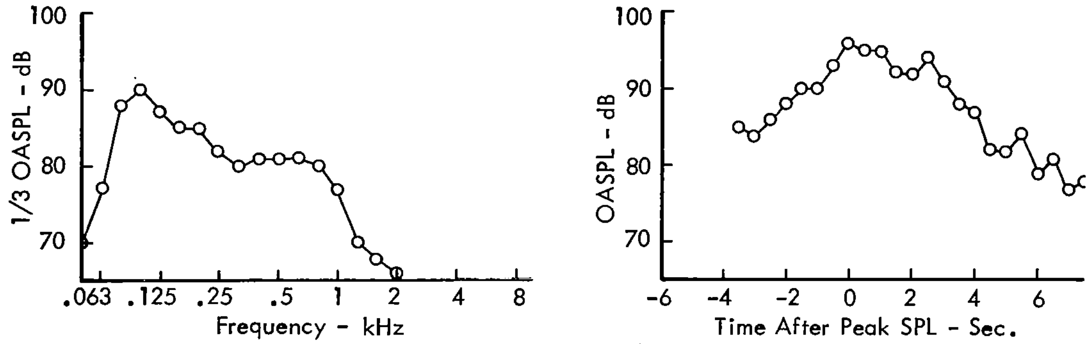


Log No. 36: Islander (Piston); Flyover; Distance - ~ 200 Ft.

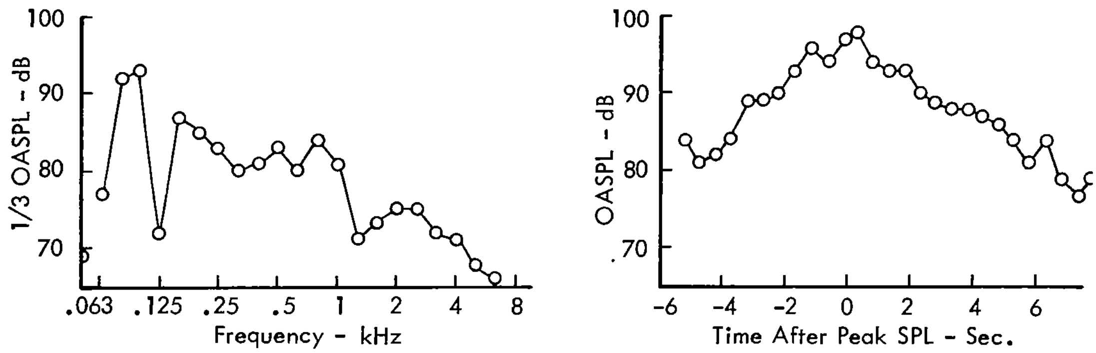
Figure 30. (Continued).



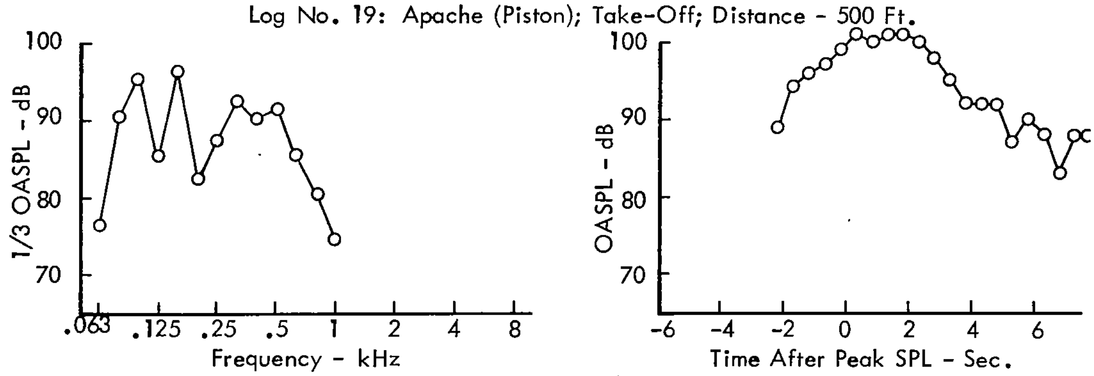
Log No. 6: Aztec (Piston); Take-Off; Distance - 300 Ft.



Log No. 16: Twin Bonanza (Piston); Take-Off; Distance - 400 Ft.

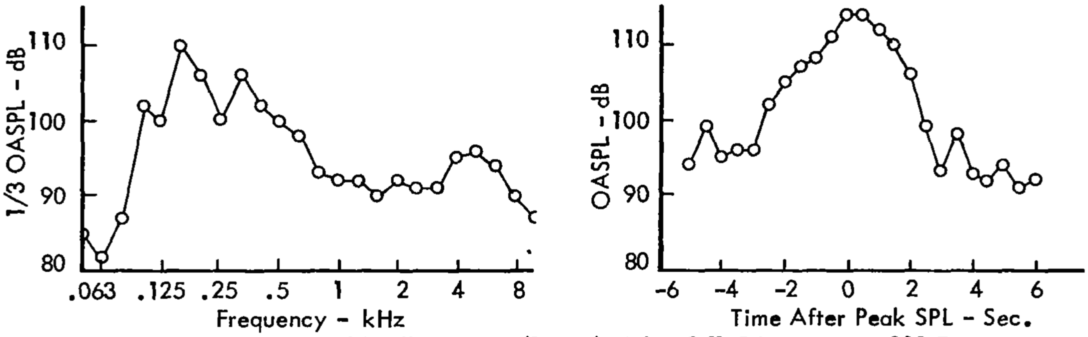


Log No. 19: Apache (Piston); Take-Off; Distance - 500 Ft.

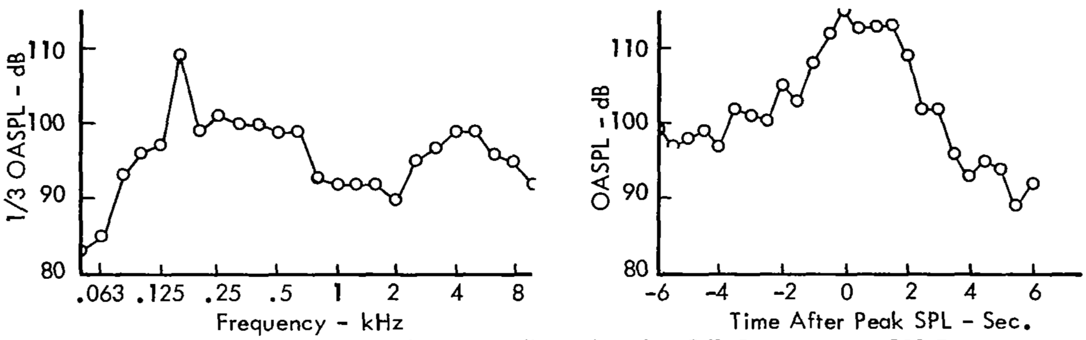


Log No. 22: Twin Cessna (Piston); Take-Off; Distance - 400 Ft.

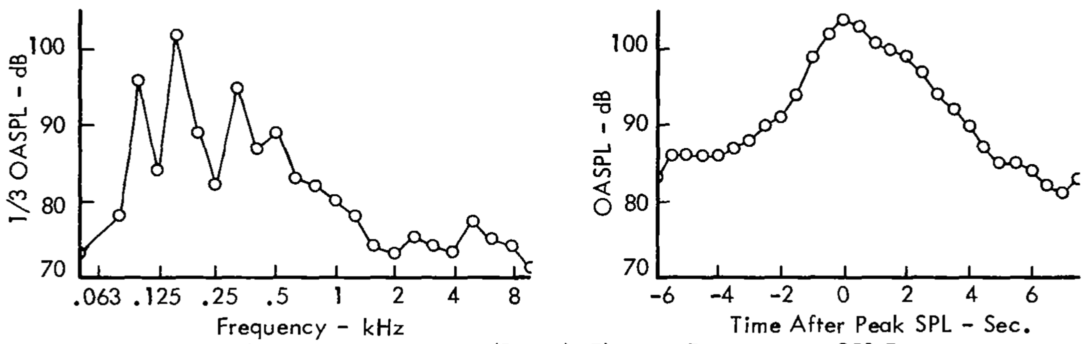
Figure 30. (Continued).



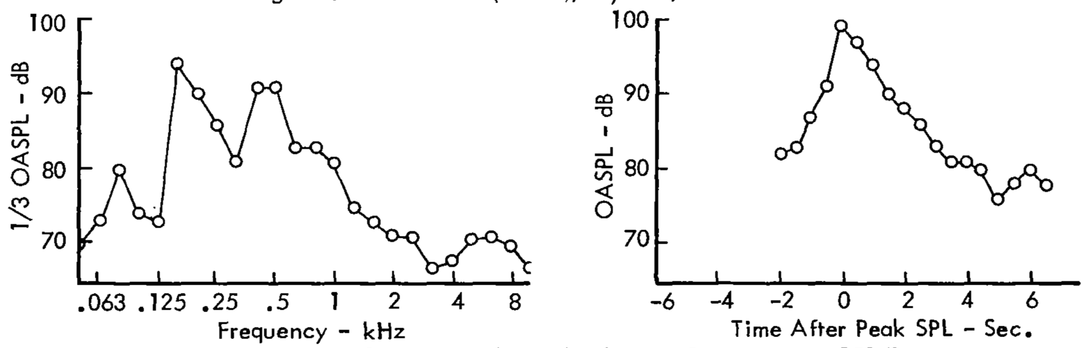
Log No. 38: Skyservant (Piston); Take-Off; Distance - ~ 200 Ft.



Log No. 39: Skyservant (Piston); Take-Off; Distance - ~ 200 Ft.

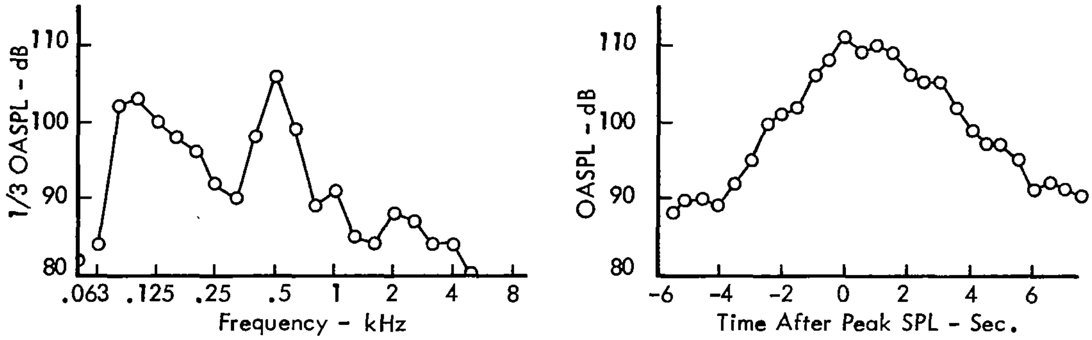


Log No. 50: Courier (Piston); Flyover; Distance - ~ 350 Ft.

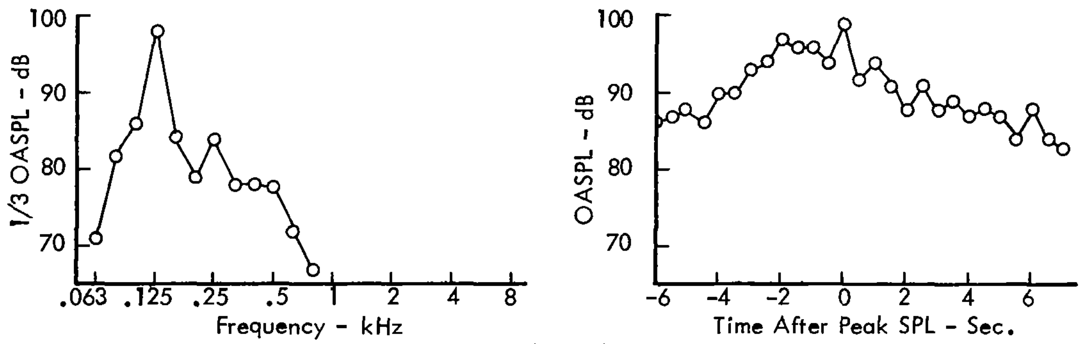


Log No. 51: Courier (Piston); Flyover; Distance - ~ 350 Ft.

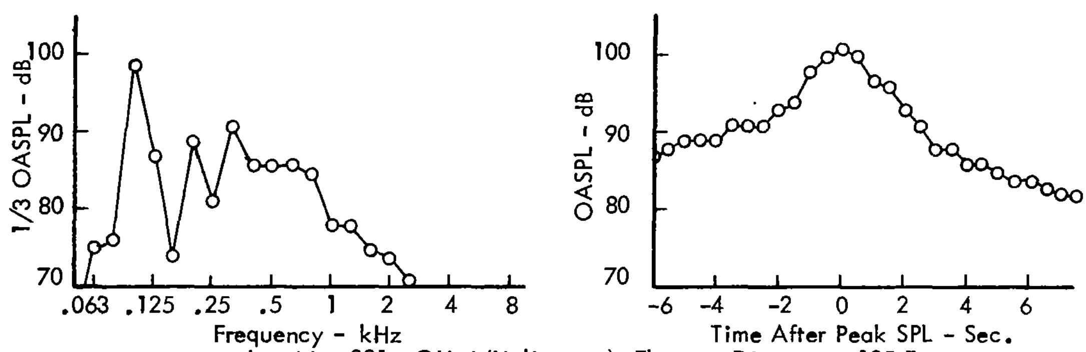
Figure 30. (Continued).



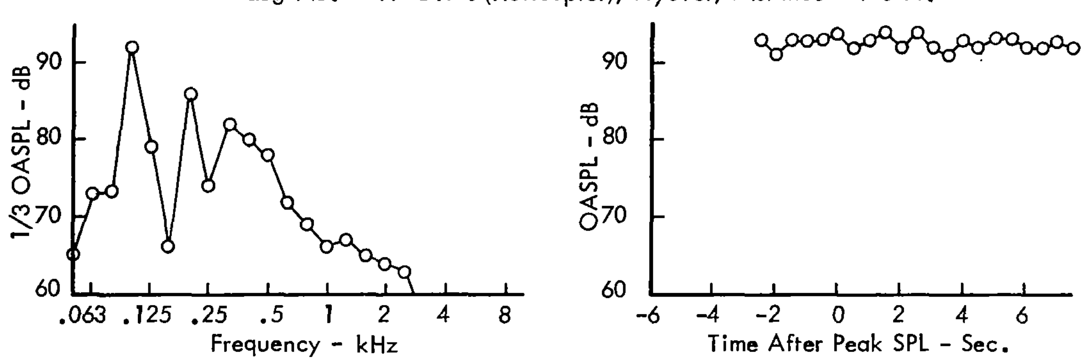
Log No. 18: Queen Air (Piston); Take-Off; Distance - 400 Ft.



Log No. 76: Queen Air (Piston); Landing; Distance - 850 Ft.

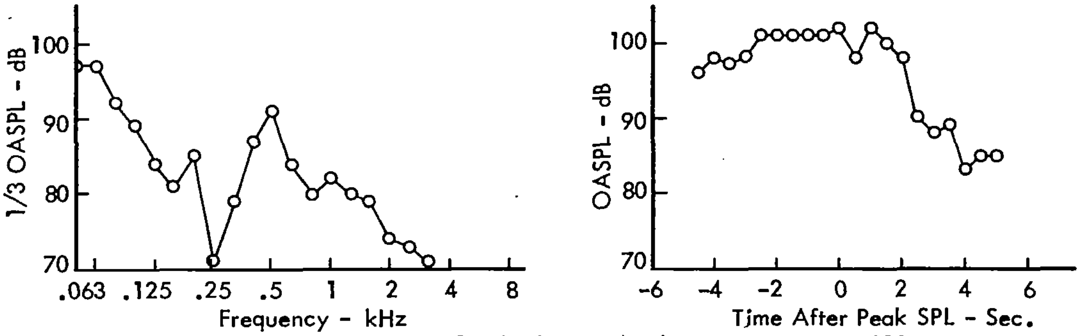


Log No. 221: OH-6 (Helicopter); Flyover; Distance - 125 Ft.

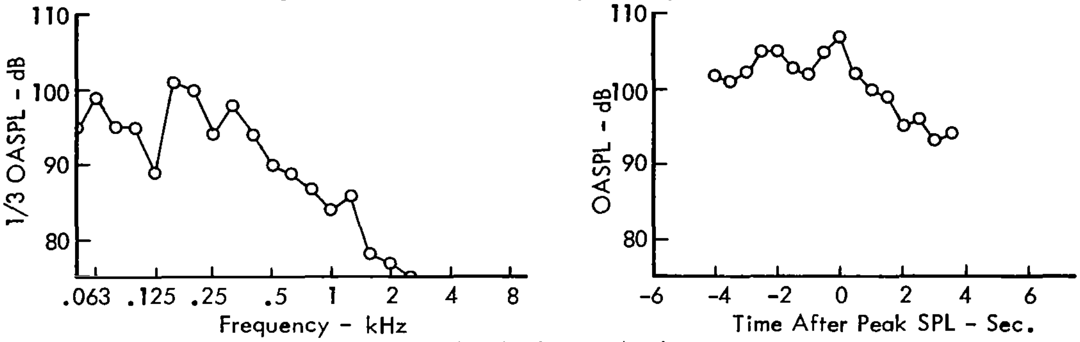


Log No. 222: OH-6 (Helicopter); Flyover; Distance - 1000 Ft.

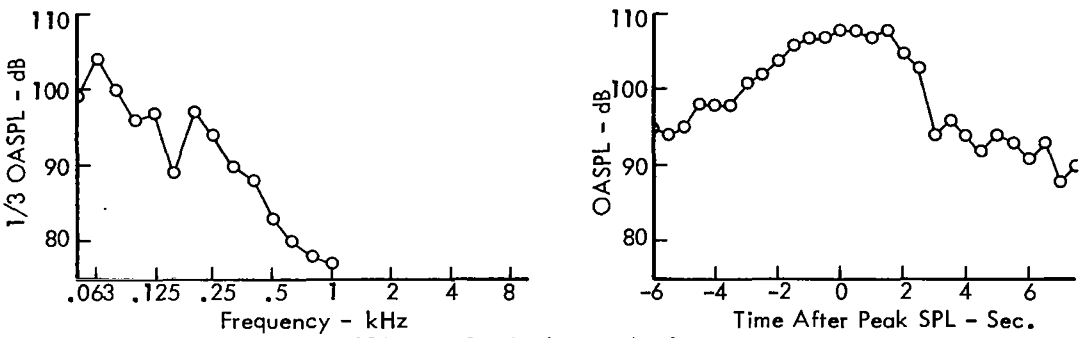
Figure 30. (Continued).



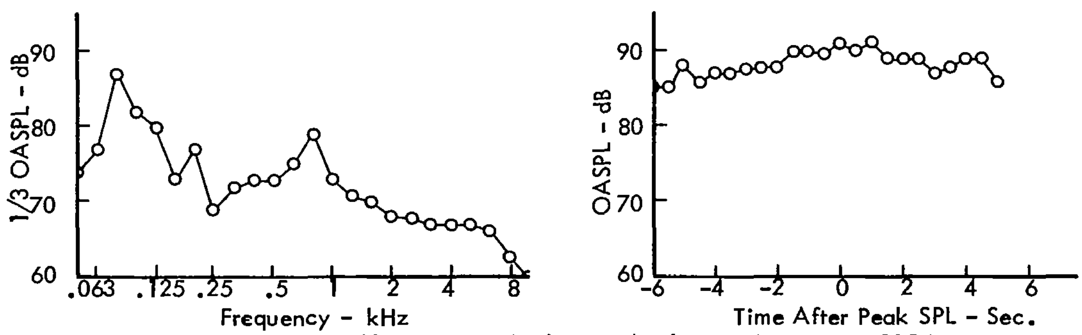
Log No. 84: AH-1G (Helicopter); Flyover; Distance - 210 Ft.



Log No. 85: AH-1G (Helicopter); Flyover; Distance - 120 Ft.

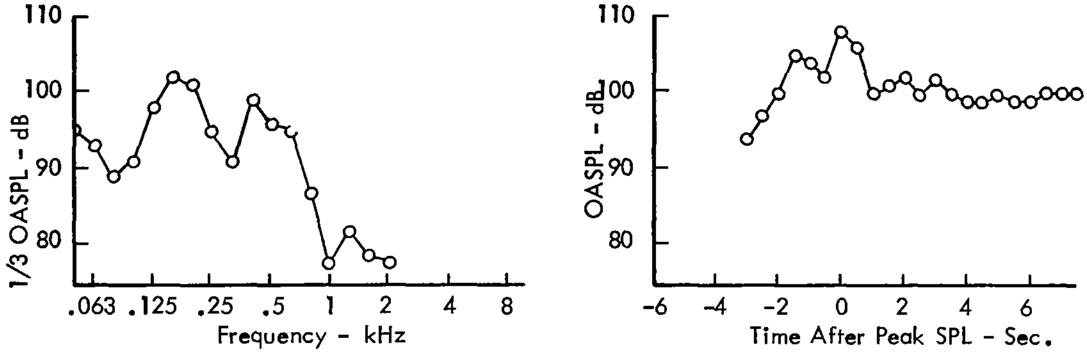


Log No. 214: AH-1G (Helicopter); Flyover; Distance - 85 Ft.

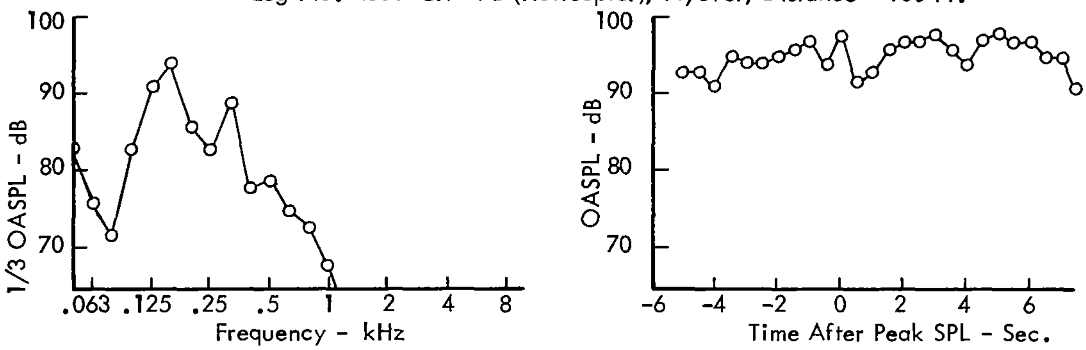


Log No. 217: QH-50 (Helicopter); Flyover; Distance - 125 Ft.

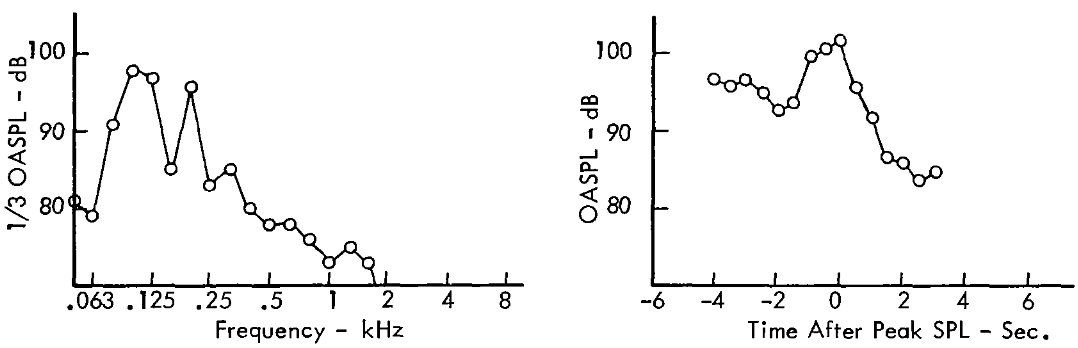
Figure 30. (Continued).



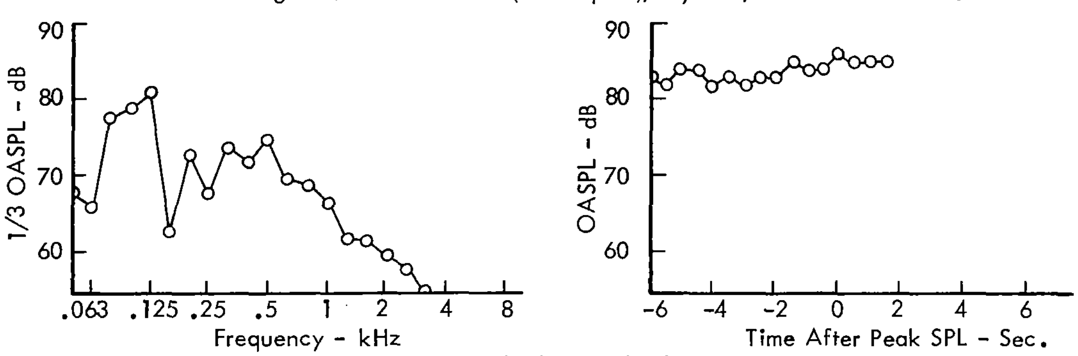
Log No. 200: CH-47B (Helicopter); Flyover; Distance - 100 Ft.



Log No. 206: CH-47B (Helicopter); Flyover; Distance - 1100 Ft.

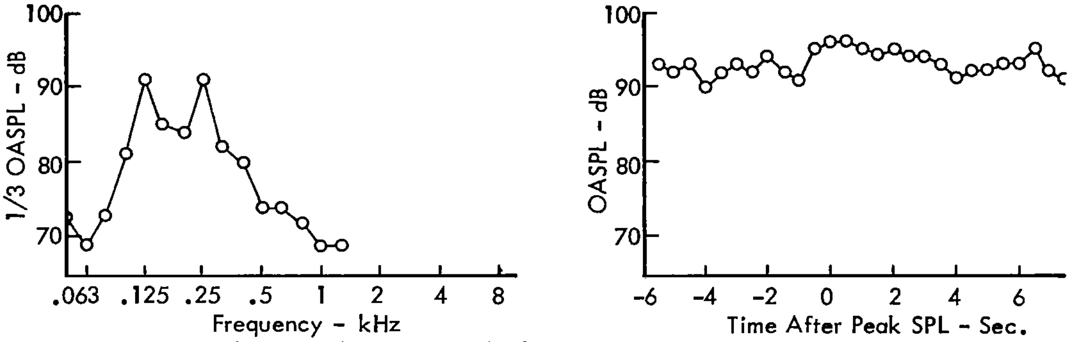


Log No. 207: CH-47B (Helicopter); Flyover; Distance - 250 Ft.

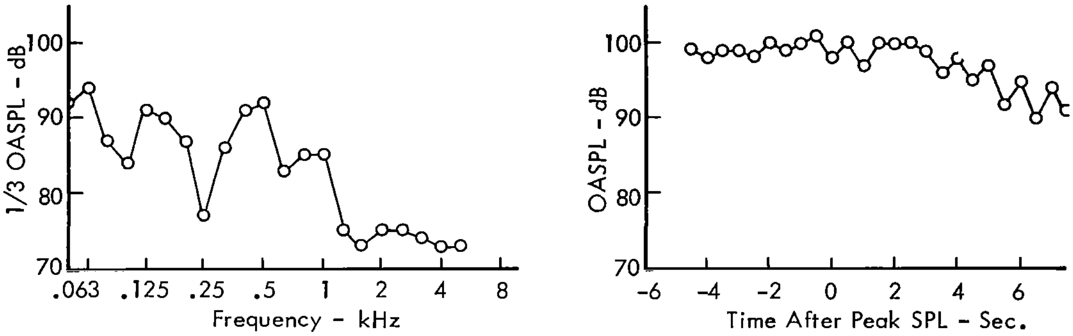


Log No. 220: QH-50 (Helicopter); Flyover; Distance - 1000 Ft.

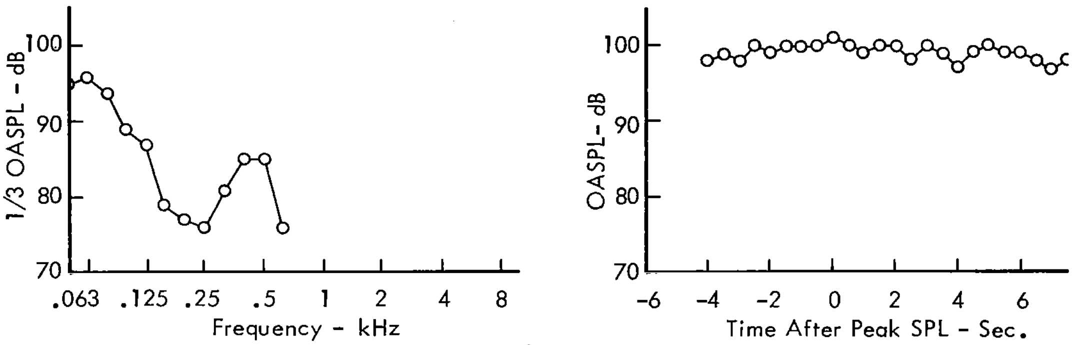
Figure 30. (Continued).



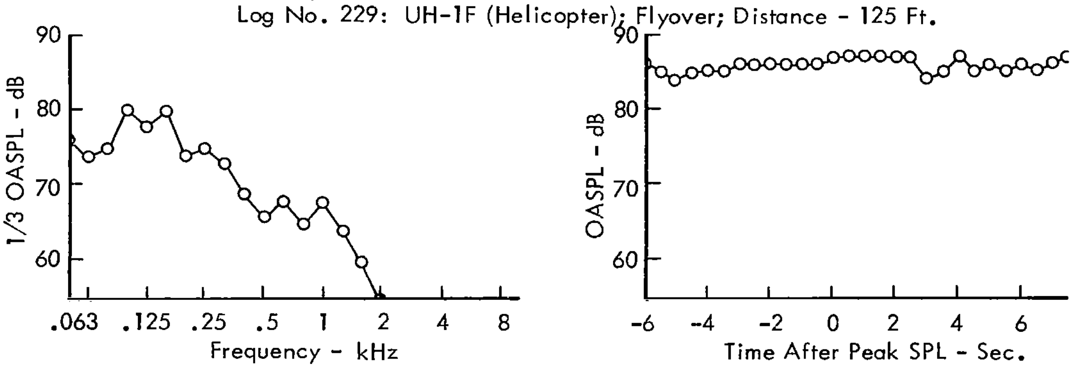
Log No. 202: UH-1B (Helicopter); Flyover; Distance - 1000 Ft.



Log No. 216: UH-1D (Helicopter); Flyover; Distance - 150 Ft.

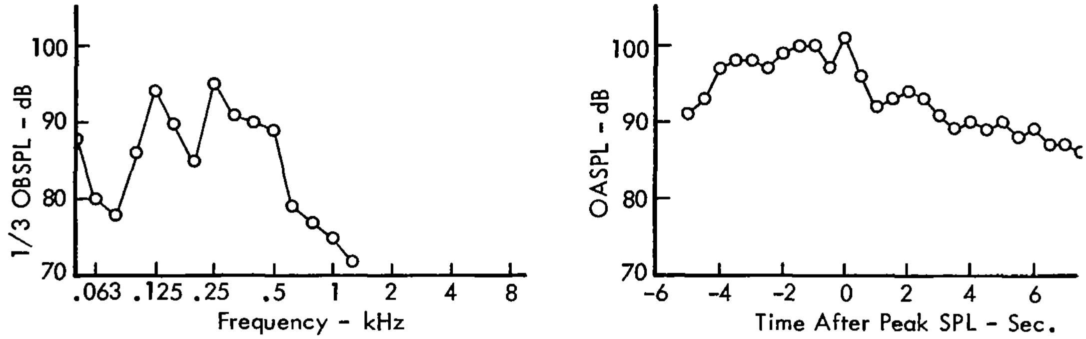


Log No. 229: UH-1F (Helicopter); Flyover; Distance - 125 Ft.

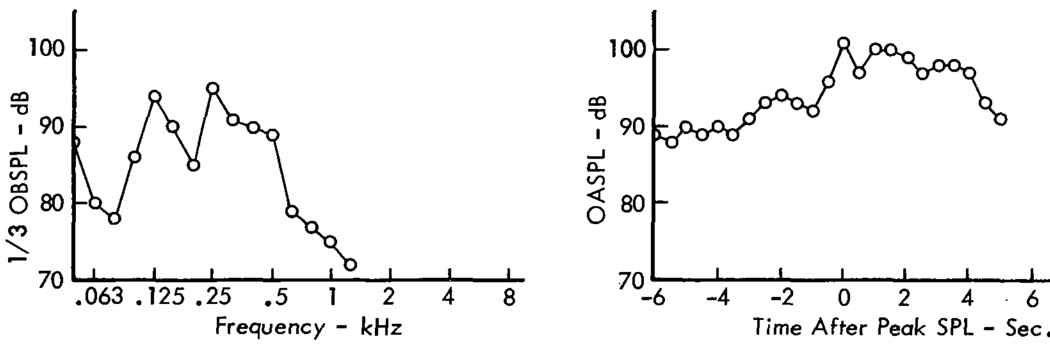


Log No. 230: UH-1F (Helicopter); Flyover; Distance - 1000 Ft.

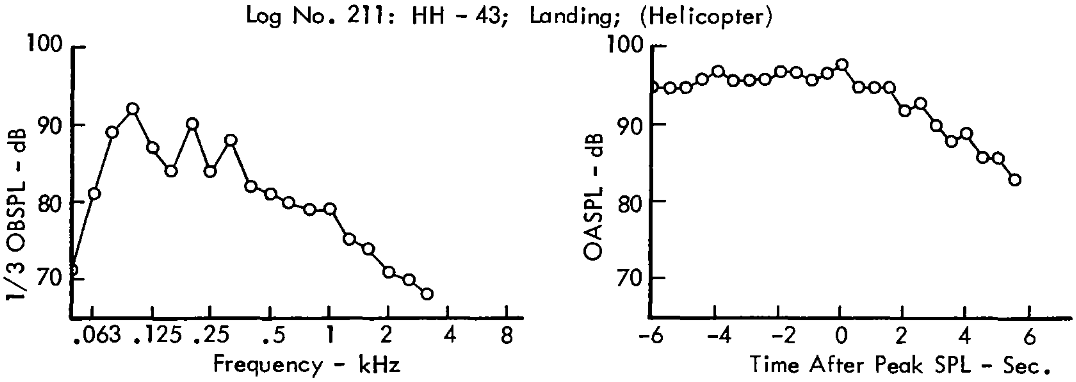
Figure 30. (Continued).



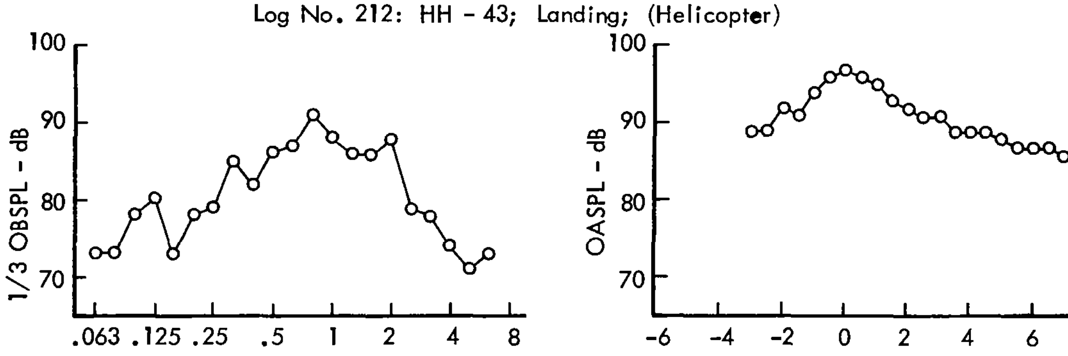
Log No. 208: HH - 43; Flyover; Distance - 200 Ft. (Helicopter)



Log No. 211: HH - 43; Landing; (Helicopter)

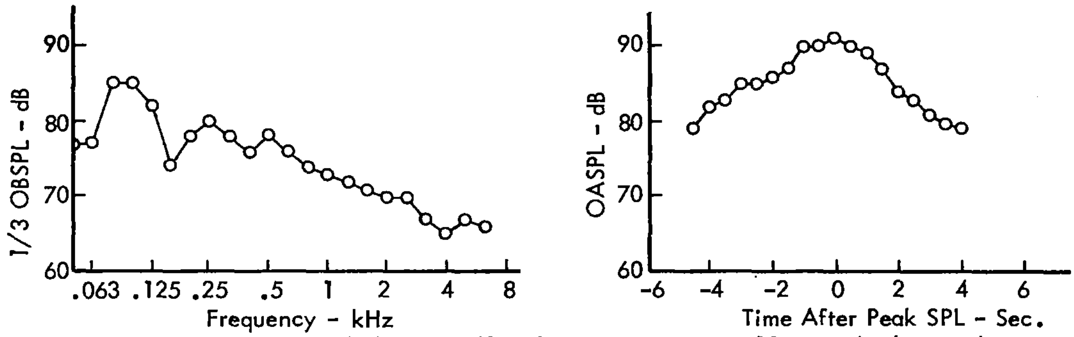


Log No. 212: HH - 43; Landing; (Helicopter)

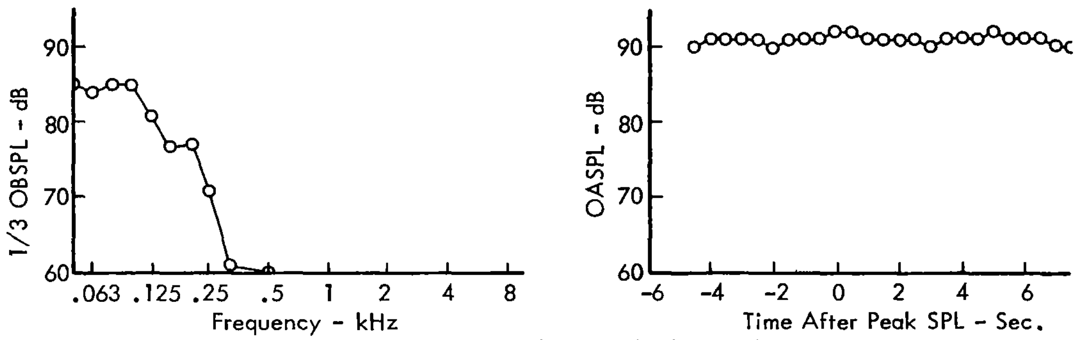


Log No. 213: HH - 43; Take - Off; (Helicopter)

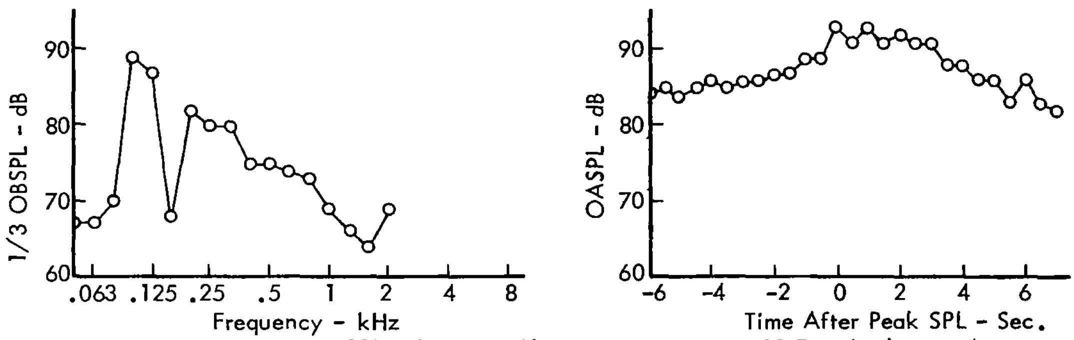
Figure 30. (Continued).



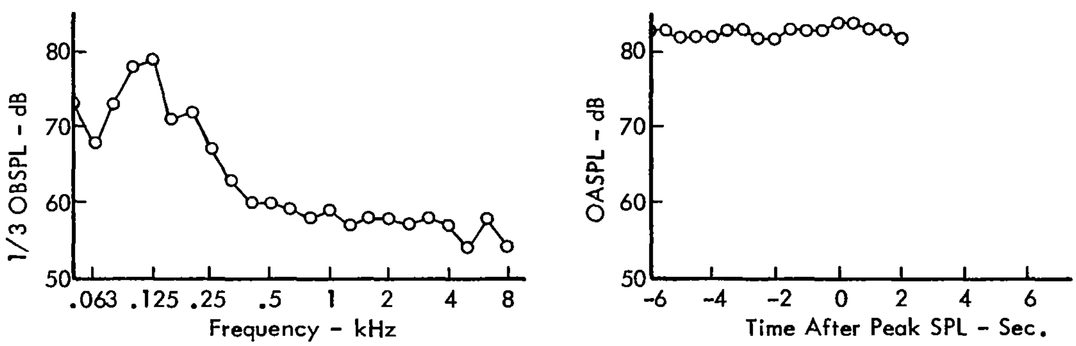
Log No. 234: HH - 43; Flyover; Distance - 125 Ft. (Helicopter)



Log No. 236: UH - 1F; Flyover; (Helicopter)

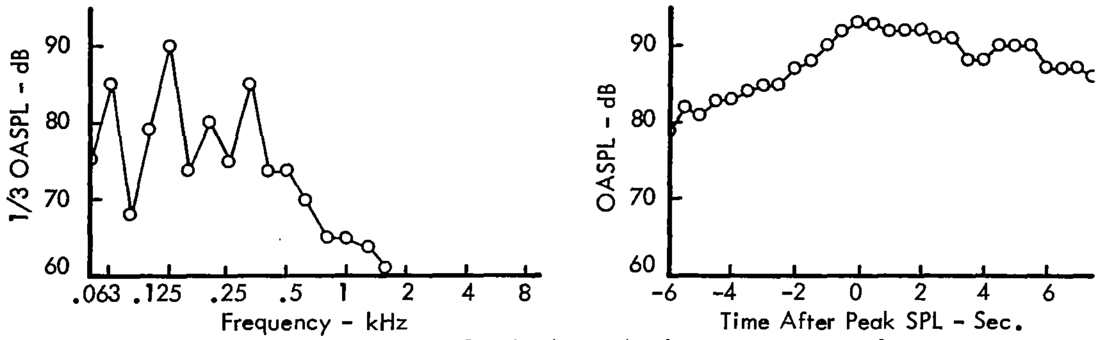


Log No. 223: OH - 6; Flyover; Distance - 500 Ft. (Helicopter)

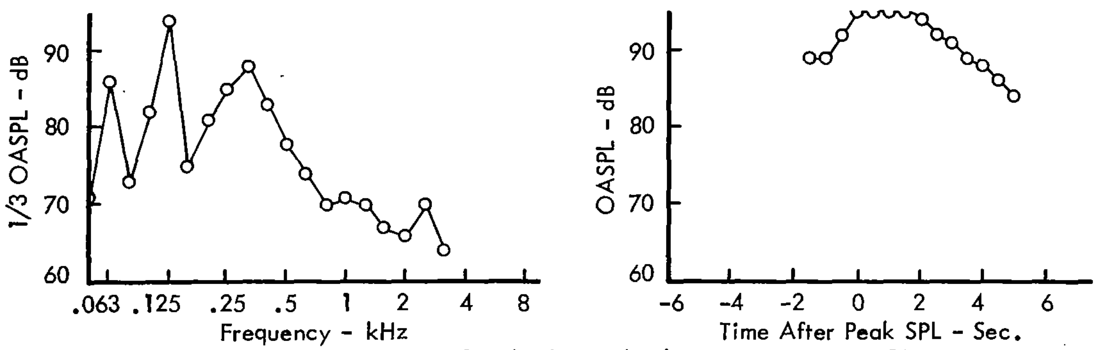


Log No. 235: HH - 43; Hover; (Helicopter)

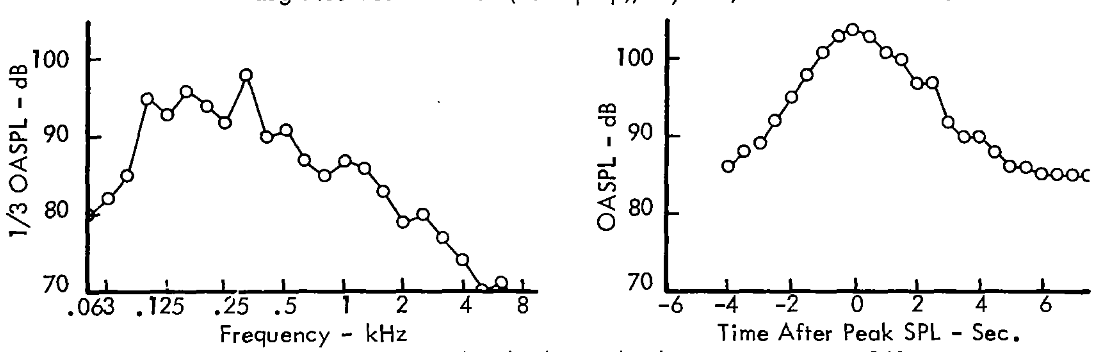
Figure 30. (Continued).



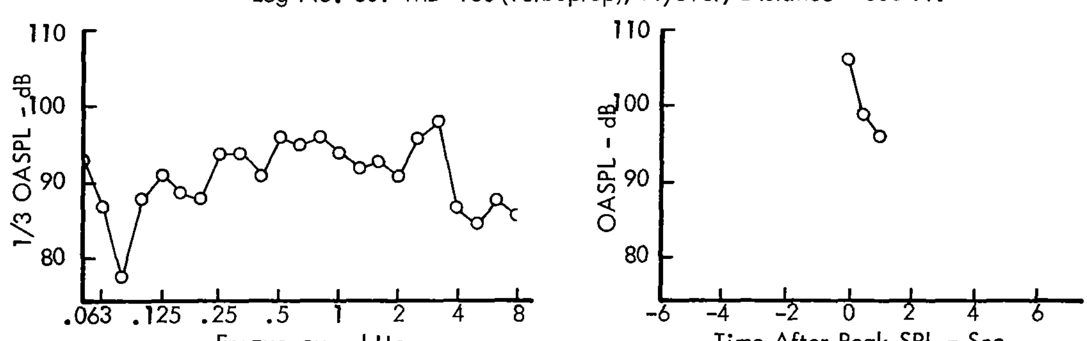
Log No. 57: MD-188 (Turboprop); Flyover; Distance - 1900 Ft.



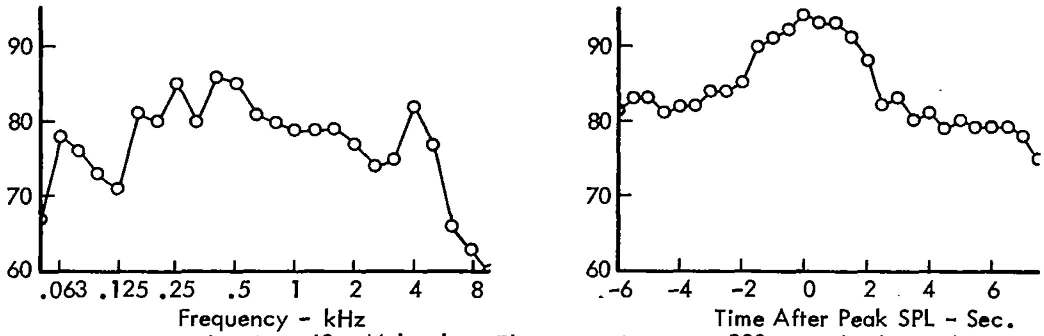
Log No. 58: MD-188 (Turboprop); Flyover; Distance - 810 Ft.



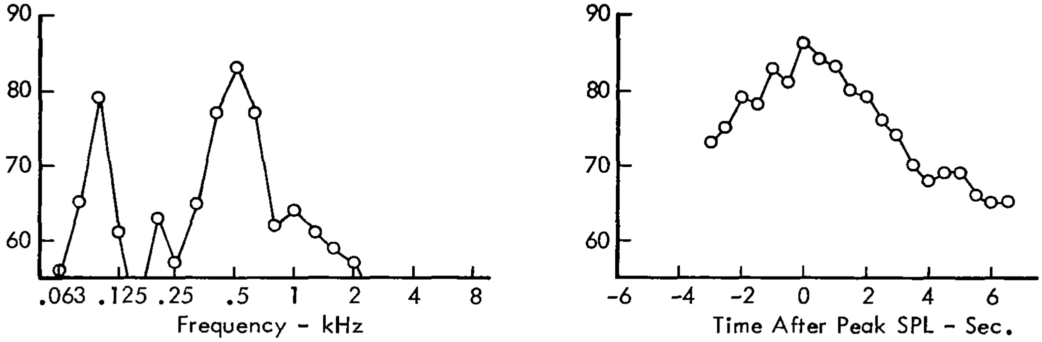
Log No. 60: MD-188 (Turboprop); Flyover; Distance - 330 Ft.



Log No. 70: Mohawk (Turboprop); Flyover; Distance - 50 Ft.



Log No. 69: Mohawk ; Flyover; Distance - 300 Ft. (Turboprop)



Log No. 71: Mohawk (Turboprop); Flyover; Distance - 300 Ft.

Figure 30. (Concluded).

NASA-Langley, 1971 — 02
Coml., Newport News, Va.